**Chemoenzymatic Synthesis of Chromodepsipeptides and Natural Product Discovery via Genome Mining**

---

**Chemoenzymatische Synthese der Chromodepsipeptide und Isolierung neuer Naturstoffe durch Genomisches Mining**

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Chemie

der Philipps-Universität Marburg

vorgelegt von

**Lars Robbel**

aus Mainz

Marburg an der Lahn, 2010

Vom Fachbereich Chemie
der Philipps-Universität Marburg als Dissertation
am 30.06.2010 angenommen.

Erstgutachter : Prof. Dr. M. A. Marahiel
(Philipps-Universität Marburg)

Zweitgutachter : Prof. Dr. L.-O. Essen
(Philipps-Universität Marburg)

Tag der Disputation: 14.07.2010

*Dedicated to my parents…*

## Abstract

Recent advances in the development of sequencing technologies have enabled the identification of a multitude of bacterial gene clusters, putatively involved in the biosynthesis of nonribosomal peptides (NRPs). Peptides of nonribosomal origin constitute a class of structurally and functionally diverse natural products, which are assembled by multimodular nonribosomal peptide synthetases (NRPSs). These compounds exhibit a broad pharmacological spectrum, ranging from antibacterial- to immunosuppressive properties. Understanding the assembly mechanisms in combination with rational genome mining approaches will provide opportunities for the discovery of new bioactive natural products.

Within this study one approach was utilized to generate thiocoraline analogs via chemoenzymatic synthesis and the second strategy focused on the *de novo* natural product discovery via genome mining.

Thiocoraline represents a pseudosymmetrical chromophore-capped octathiodepsipeptide, in which the symmetrical halves are linked via thioester bonds. In this study, the cyclodimerization potential of the thioesterase domain of the thiocoraline biosynthetic machinery (TioS PCP-TE) was investigated to obtain further insights into the iterative assembly of chromodepsipeptides. To address this objective, the recombinant enzyme was incubated with synthetically derived tetrapeptidyl substrates, resembling thiocoraline precursors. It was shown that the enzyme catalyzes the cyclodimerization of linear precursor molecules and an unprecedented macrothiolactonization. Evaluation of the biocombinatorial potential established the thioesterase as a robust and versatile catalyst for the generation of chromodepsipeptide analogs, harbouring thioester- or ester-linkages. As thiocoraline attains its antitumor activity from DNA-bisintercalation, the chemoenzymatically generated macrocycles were isolated and investigated towards DNA-bisintercalation activity *in vitro*.

In the second part of this study, bioinformatic analysis of the 8.2 Mb *Saccharopolyspora erythraea* genome revealed two cryptic NRPS gene clusters related to hydroxamate-type siderophore biosynthesis. Detailed analysis of adenylation domain substrate-specificity and module organization enabled the establishment of a highly selective and sensitive radio-LCMS-guided genome mining approach. Application of this approach resulted in the discovery of the siderophore erythrochelin. Structure elucidation of erythrochelin was accomplished via NMR- and $MS^n$-analysis and revealed the sequence of the tetrapeptide siderophore to be: α-*N*-acetyl-δ-*N*-acetyl-δ-*N*-hydroxy-D-ornithine-D-serine-cyclo(δ-*N*-hydroxy-L-ornithine-δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine).

Erythrochelin assembly requires the proliferation of δ-*N*-hydroxy-L-ornithine (L-hOrn) and δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine (L-haOrn). The corresponding modifying enzymes, the FAD-dependent monooxygenases EtcB and Sace_1309 together with the bifunctional malonyl-CoA decarboxylase/*N*-acetyltransferase were identified and biochemically characterized. *In vitro* studies revealed EtcB and Sace_1309 to exclusively catalyze the δ-*N*-hydroxylation of free L-ornithine. The second tailoring enzyme, Mcd, was shown to catalyze malonyl-CoA decarboxylation and subsequent acetyltransfer onto the δ-hydroxamino group of L-hOrn, affording L-haOrn. Based on the elucidation of precursor biosynthesis (L-haOrn), a model for the entire erythrochelin assembly is presented.

## Zusammenfassung

Die Entwicklung neuartiger Sequenzierungstechnologien ermöglicht die Identifizierung einer Vielzahl von kryptischen bakteriellen Genclustern, welche an der Biosynthese nichtribosomaler Peptide (NRPs) beteiligt sind. Peptide nichtribosomalen Ursprungs konstituieren eine Klasse strukturell diverser Naturstoffe, welche durch multimodulare Peptidsynthetasen (NRPSs) assembliert werden. Nichtribosomale Peptide weisen ein breites pharmakologisches Spektrum als Antitumor-Wirkstoffe, Antibiotika oder Immunosuppressiva auf. Das eingehende Verständnis der nichtribosomalen Assemblierungsmechanismen, in Kombination mit rationalem genomischen Mining ermöglicht die Identifizierung neuartiger, bioaktiver Verbindungen.

Die vorliegende Arbeit umfasst neben der chemoenzymatischen Synthese von Thiocoralin-Analoga die *de novo* Identifizierung neuer Naturstoffe durch genomisches Mining.

Thiocoralin stellt ein makrozyklisches Oktathiodepsipeptid dar, in welchem die mit exozyklischen Chromophoren funktionalisierten Oligopeptide über zwei Thioesterbindungen verbunden sind. In dieser Arbeit wurde das Zyklodimerisierungspotential der Thiocoralin Thioesterase-Domäne (TioS PCP-TE) untersucht, um ein genaueres Verständnis der iterativen Assemblierung der Chinolin- und Chinoxalinchromodepsipeptide zu erhalten. Hierzu wurde das Enzym rekombinant produziert und mit synthetischen, linearen Peptidylsubstratanaloga inkubiert. TioS PCP-TE katalysierte die Zyklodimerisierung der linearen Tetrapeptidylsubstrate unter Ausbildung verschiedener Makrothiolaktone. Die Evaluierung des biokombinatorischen Potentials ergab eine relaxierte Substratspezifität des Enzyms und ermöglichte die Generierung von Chromodepsipeptid-Analoga. Die chemoenzymatisch generierten makrozyklischen Laktone oder Thiolaktone wurden isoliert und auf ihre DNA-Bisinterkalationsaktivität hin untersucht.

Im zweiten Teil dieser Arbeit resultierte die bioinformatische Analyse des 8.2 Mb umfassenden Genoms von *Saccharopolyspora erythraea* in der Identifizierung von zwei kryptischen Hydroxamat-Siderophor NRPS Genclustern. Die Analyse der Adenylierungs-Domänen Substratspezifität und der modularen Organisation ermöglichte die Etablierung eines hochselektiven und hochsensitiven Radio-LCMS-geleiteten genomischen Mining Ansatzes. Die Anwendung dieses Ansatzes resultierte in der Entdeckung des Siderophors Erythrochelin. Die strukturelle Charakterisierung von Erythrochelin wurde mittels NMR-spektroskopischen sowie massenspektrometrischen Studien realisiert und ergab für die Sequenz des Tetrapeptid-Siderophors: α-*N*-Acetyl-δ-*N*-acetyl-δ-*N*-hydroxy-D-Ornithin-D-Serin-zyklo(δ-*N*-Hydroxy-L-Ornithin-δ-*N*-Acetyl-δ-*N*-hydroxy-L-Ornithin).

Die Assemblierung von Erythrochelin erfordert die Synthese von δ-*N*-Hydroxy-L-Ornithin (L-hOrn) und δ-*N*-Acetyl-δ-*N*-hydroxy-L-Ornithin (L-haOrn). Zur Untersuchung der korrespondierenden Biosynthesen wurden die FAD-abhängigen Monooxygenasen EtcB und Sace_1309, sowie die bifunktionelle Malonyl-CoA Decarboxylase/*N*-Acetyltransferase Mcd, identifiziert und biochemisch charakterisiert. Mittels *in vitro*-Studien wurde gezeigt, dass EtcB und Sace_1309 ausschließlich die δ-*N*-Hydroxylierung von freiem L-Ornithin katalysieren. Mcd katalysierte die Decarboxylierung von Malonyl-CoA und einen anschließenden Acetyltransfer auf die δ-Hydroxamino Funktionalität von L-hOrn unter Bildung von L-haOrn. Die Aufklärung der Biosynthese des Vorläufermoleküls L-haOrn resultierte in der Postulierung eines umfassenden Modells für die Assemblierung von Erythrochelin.

***The majority of the work presented herein has been published:***

**Robbel**, **L**., Hoyer, K.M., Marahiel, M.A., (2009), TioS T-TE - a prototypical thioesterase responsible for the cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides, *FEBS J.*, 276, 1641-1653.

**Robbel**, **L**., Knappe, T.A., Linne, U., Xie, X., Marahiel, M.A., (2010), Erythrochelin - a hydroxamate-type siderophore predicted from the genome of *Saccharopolyspora erythraea*, *FEBS J.*, 277, 663-676.

***Additional publications:***

Knappe, T.A., Linne, U., **Robbel**, **L**., Marahiel, M.A., (2009), Insights into the biosynthesis and stability of the lasso peptide capistruin, *Chem. Biol.*, 16, 1290-1298.

**Robbel**, **L**., Marahiel, M.A., (2010), Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery, *J. Biol. Chem*., in press, doi: 10.1074/jbc.R110.128181.

## Abbreviations

| | |
|---|---|
| 3HQA | 3-hydroxy-quinaldic acid |
| A-domain | adenylation domain |
| aa | amino acid |
| ACE | angiotensin converting enzyme |
| ac-haOrn | α-*N*-acetly-δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine |
| ACP | acyl carrier protein |
| ACT | Artemis Comparison Tool |
| α-KG | α-ketoglutarate |
| AMP | adenosine monophosphate |
| AT | acetyltransferase |
| ATP | adenosine-5'-triphosphate |
| BLAST | Basic Local Alignment Search Tool |
| Boc | *tert*-butoxycarbonyl |
| bp | base pairs |
| C-domain | condensation domain |
| Cy-domain | heterocyclization domain |
| CAS | chromazurol S |
| CDA | calcium-dependent antibiotic |
| CDS | coding sequence |
| CoA | coenzyme A |
| COSY | correlation spectroscopy |
| DA | dalton |
| DCM | dichloromethane |
| DHB | 2,3-dihydroxybenzoic acid |
| DIPEA | *N*,*N*-diisopropylethylamine |
| DKP | diketopiperazine |
| DMF | dimethyl formamide |
| DMSO | dimethylsulfoxide |
| DNA | deoxyribonucleic acid |
| DQF | double quantum filtering |
| E-domain | epimerization domain |
| EDA | ethylenediamine |
| EDC | *N*-(3-dimethylaminopropyl)-*N*'-ethylene carbodiimide hydrochloride |
| EDTA | ethylenediaminetetraacetic acid |
| EIC | extracted ion chromatogram |
| ESI | electron-spray ionization |
| F-domain | *N*-formylation domain |
| Fl-CoA | fluoresceinyl-CoA |
| FT-ICR | fourier transform ion cyclotron resonance |
| FP-TAMRA | fluorophosphonate-*N,N,N´,N´*-tetramethylrhodamine |
| FA | fatty acid |
| FAD | flavin adenine dinucleotide |
| FDAA | [*N*-α-(2,4-dinitro-5-fluorophenyl)-L-alaninamide] |
| FMOC | fluorenylmethyloxycarbonyl |
| FPLC | fast protein liquid chromatography |
| haOrn | δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine |

| | |
|---|---|
| HBTU | *O*-Benzotriazole-*N*,*N*,*N*',*N*'-tetramethyl-uronium-hexafluoro-phosphate |
| HEPES | 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid |
| HHL | *N*-hippuryl-L-histidyl-L-leucine |
| Hip | hippuric acid |
| HL | L-histidyl-L-leucine |
| HMBC | heteronuclear multiple bond coherence |
| HOAt | 1-hydroxy-7-azabenzotriazol |
| HOBt | hydroxybenzotriazole |
| hOrn | δ-*N*-hydroxy-L-ornithine |
| HPLC | high performance liquid chromatography |
| HRMS | high resolution mass spectrometry |
| HSQC | heteronuclear single-quantum correlation spectroscopy |
| IPTG | isopropyl-β-D-thiogalactopyranoside |
| kbp | kilo base pairs |
| kDa | kilo Dalton |
| lac | lactose |
| MT-domain | methyltransferase domain |
| mbp | mega base pairs |
| MDa | mega Dalton |
| MeCN | acetonitrile |
| MIC | minimum inhibitory concentration |
| MOPS | 3-(*N*-morpholino)propanesulfonic acid |
| mRNA | messenger ribonucleic acid |
| MRSA | methicillin-resistant *Staphylococcus aureus* |
| MS | mass spectrometry |
| NDP | nucleoside diphosphate |
| NIS | NRPS independent siderophore |
| NOE | nuclear Overhauser effect |
| NRP | nonribosomal peptide |
| NRPS | nonribosomal peptide synthetase |
| NTA | nitrilotriacetic acid |
| Ox-domain | oxidation domain |
| OD | optical density |
| ORF | open reading frame |
| PCP | peptidyl-carrier-protein |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| PK | polyketide |
| PKS | polyketide synthase |
| ppan | 4'-phosphopantetheine |
| PPAT | phosphopantetheine adenosine transferase |
| PPi | inorganic pyrophosphate |
| PPTase | 4'-phosphopantetheine transferase |
| PyBop | benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate |
| QA | quinaldic acid |
| QTOF | quadrupole time-of-flight |
| QX | quinoxaline-2-carboxylic acid |
| R-domain | reductase domain |

| | |
|---|---|
| RNA | ribonucleic acid |
| RP | reversed-phase |
| RT | room temperature |
| SAM | *S*-adenosylmethionine |
| SDS | sodium dodecyl sulfate |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SIC | selected ion chromatogram |
| SNAC | *N*-acetylcysteamine |
| sp. | species |
| str. | strain |
| TCEP | tris-(carboxyethyl)-phosphine |
| TE | thioesterase domain |
| TEII | type II thioesterase |
| TES | triethyl silane |
| TFA | trifluoroacetic acid |
| TFE | trifluoroethanol |
| TIC | total ion chromatogram |
| TIPS | triisopropylsilane |
| TOCSY | total correlation spectroscopy |
| $t_R$ | retention time in minutes |
| TRIS | tris-(hydroxymethyl)-aminomethane |
| Trt | trityl |
| tRNA | transfer ribonucleic acid |
| VRE | vancomycin-resistant Enterococci |
| w/o | without |

**Table 1:** Overview of the proteinogenic amino acids. The three- and one-letter codes are given for each amino acid as well as the molecular weight.

| amino acid | three letter code | one letter code | MW (Da) |
|---|---|---|---|
| alanine | Ala | A | 89 |
| arginine | Arg | R | 174 |
| asparagine | Asn | N | 132 |
| aspartic acid | Asp | D | 133 |
| cysteine | Cys | C | 121 |
| glutamic acid | Glu | E | 147 |
| glutamine | Gln | Q | 146 |
| glycine | Gly | G | 75 |
| histidine | His | H | 155 |
| isoleucine | Ile | I | 131 |
| leucine | Leu | L | 131 |
| lysine | Lys | K | 146 |
| methionine | Met | M | 149 |
| phenylalanine | Phe | F | 165 |
| proline | Pro | P | 115 |
| serine | Ser | S | 105 |
| threonine | Thr | T | 119 |
| tryptophan | Trp | W | 204 |
| tyrosine | Tyr | Y | 181 |
| valine | Val | V | 117 |

## 1. Introduction

The discovery of natural products as biologically active drugs had a drastic impact on fatality rates and gave rise to a variety of drug classes for the treatment of a multitude of diseases.[1] Compounds derived from the secondary metabolism of bacteria and fungi therefore represent a rich resource of pharmacologically relevant natural products.[2]

It is assumed that immobile microorganisms which do not possess the evolutionary potential to develop physical defense mechansims are the most promising candidates for the discovery of bioactive secondary metabolites.[3] The selection pressure these microorganisms are subjected to led to the directed evolution of natural products that are optimized for chemical defense or inter- or intra-species communication and convey evolutionary advantages over competing organisms in the natural habitat. Compounds targeting enzymes, receptors or membranes, which exhibit highly conserved structural or chemical properties, are potential candidates for the development of novel therapeutics or therapies.[4] The compounds isolated from prokaryotes or eukaryotes comprise peptides, polyketides, steroids, glycosphingolipids and terpenes exhibiting a broad pharmacological and chemical spectrum.[5] Among those compounds the class of biologically active peptides represents a rich resource for the discovery of novel bioactive agents. The biosynthesis of the oligopeptides can be either carried out via a ribosomal strategy, as in the case of capistruin and patellamide or via a template-directed manner by multimodular nonribosomal peptide synthetases (NRPSs).[6-8] Peptides of nonribosomal origin display a remarkable structural and functional diversity arising from biosynthetic mechanisms employed in the assembly of the oligopeptides as well as from the over 500 currently known building blocks (Figure 1.1). Nonribosomal peptides (NRPs) exhibit a broad range of biological activities ranging from antitumor- (bleomycin), antibacterial- (gramicidin S) or immunosuppressive-activities (cyclosporine).[9-11] The US Food and Drug Administration (FDA)-approval of numerous NRPS-derived compounds confirms the pharmacological potential of this intriguing peptide class (Table S1- Supplementary Section).

A main drawback of the extensive treatment of bacterial infections is the rapid bacterial acquisition of resistance to conventional antibiotics, representing an increasing challenge in treating infections with the contemporary drug-arsenal. Especially the rise of drug-resistant Gram-positive pathogens, exemplified by methicillin-resistant *Staphylococcus aureus* (MRSA) or vancomycin-resistant Enterococci (VRE), underlines the urgent demand for NRPS-derived antibiotics with alternative modes of action.[12] Although the effort for the identification of new bioactive NRPs has drastically increased in the past

decades, only three new antibacterial classes have been approved by the Food and Drug Administration (FDA) since the 1970s.[13]

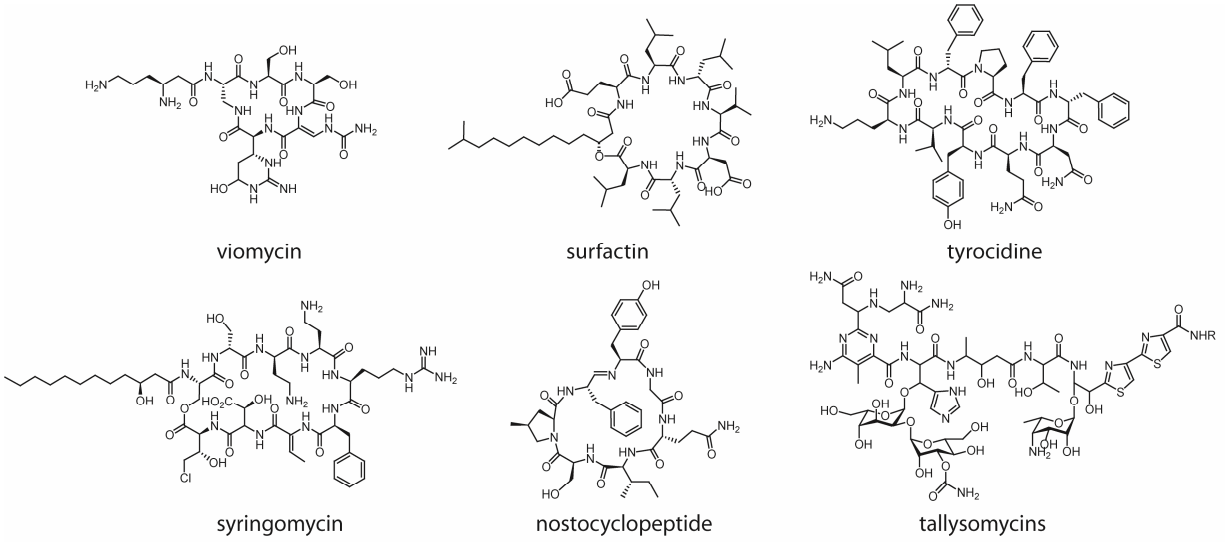



**Figure 1.1:** Selection of nonribosomally assembled and biologically active natural products. The structural diversity is a result of the biosynthetic mechanism of NRP assembly and the incorporation of over 500 unique building blocks. A structural key feature is the macrocyclic core of the NRP, conveying protection against proteolytic degradation. NRPs shown are: viomycin (*Streptomyces* sp. strain ATCC 11861), surfactin (*Bacillus subtilis*), tyrocidine (*Bacillus brevis*), syringomycin (*Pseudomonas syringaea*), nostocyclopeptide (*Nostoc* sp. ATCC 53789) and the tallysomycins (*Streptoalloteichus hidustanus* E465-94 ATCC 31158).[14-19]

In the course of this work two strategies were applied for the enhancement of structural and functional diversity among nonribosomal peptides. The first part describes the chemoenzymatic synthesis of derivatives of thiocoraline (*Micromonospora* ML-1), an antitumor compound currently undergoing preclinical trials phase II.[20] The second part focuses on the discovery of natural products via rational genome mining strategies, resulting in the detection of the hydroxamate-type siderophore erythrochelin (*Saccharopolyspora erythraea* NRRL 23338).[21] Understanding the ribosome-independent assembly of biologically active natural products in combination with rational genome mining approaches will provide opportunities for the generation of NRP-derivatives or lead to the discovery of novel natural products for potential clinical applications.

### 1.1 The Nonribosomal Assembly of Peptides

In contrast to the ribosomal assembly of peptides in which the mRNA serves as a template, the NRPS represents template and biosynthetic machinery at once. Genetic and biochemical characterization of the NRPS enzymes revealed a multiple-carrier-thiotemplate mechanism for oligopeptide assembly.[22-24] According to this model, the NRPSs represent megaenzymes which can be subdivided into modules and domains.[25] Each module contains all catalytic units for the specific recognition, activation, covalent binding and incorporation of a building block into the oligopeptide chain and can be dissected into individual domains.[26] The identification of highly conserved core motifs within the specific domains allows the definition of domain borders and facilitates primary sequence prediction of the product. In NRPS-systems following a linear logic of oligopeptide assembly, the number of building blocks found in the product directly correlates with the number of modules. Exceptions to this assembly mechanism are discussed in detail (Chapter 1.4).[27] A minimal module consists of the essential domains adenylation (A)-domain, condensation (C)-domain and peptidyl-carrier-protein (PCP) and contains all catalytic units required for peptide-bond formation and translocalization of peptidyl intermediates. Recently, the three-dimensional structure of an entire NRPS-module was elucidated, granting insights into unique inter-domain communications.[28] The C-domain and the $A_{core}$-domain constitute a catalytic platform on which the PCP and the *C*-terminal $A_{sub}$-domain are located. In addition, the crystal structure revealed the thioesterase (TE)-domain, responsible for macrocyclization or hydrolytic product release, to be associated with the PCP via a short linker region (Figure 1.2).

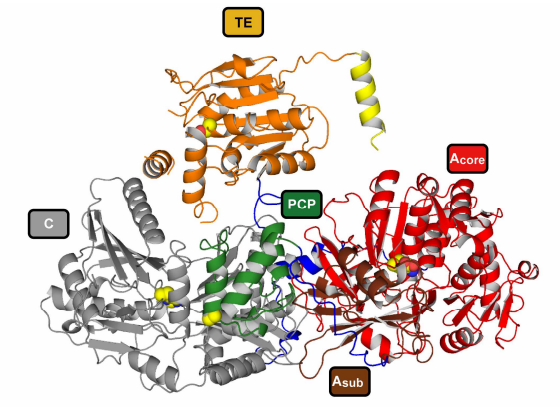



**Figure 1.2:** The three-dimensional structure of the SrfA-C termination module (C-A-PCP-TE) involved in surfactin biosynthesis (PDB code: 2VSQ). The C-domain (grey) with the active-site His147-residue is connected to the $A_{core}$-domain (red) in which a Leu-residue is bound to the active-site. The *N*-terminal $A_{sub}$-domain (brown) is connected to the PCP (green) in which the serine of the core-T motif is substituted with alanine. The TE-domain (orange) is linked to the PCP via a short linker-region. Linker regions are accentuated in blue.

The PCP has to be posttranslationally modified in order to serve as the universal transporter unit of the assembly line. This modification is catalyzed by phosphopantetheinyl-transferases (PPtases), converting the *apo*-PCP to *holo*-PCP by transfer of the phosphopantetheine-cofactor (ppan) to a highly conserved serine residue.[29]

*In cis* acting modifying domains which are integrated into the NRPS or tailoring enzymes operating *in trans* contribute to the vast structural diversity of NRPs.[30-31] A mechanistic overview of peptide bond formation catalyzed by a minimal NRPS-system is given in Figure 1.3.

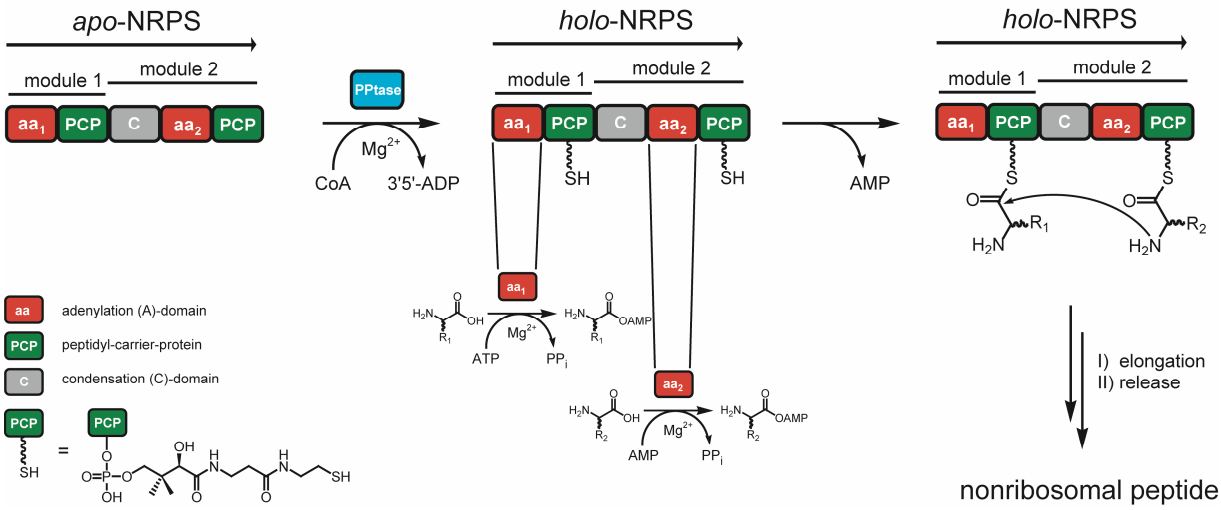



**Figure 1.3:** Mechanistic overview of NRPS-catalyzed peptide bond formation by the essential domains of the synthetase. Conversion of the *apo*-NRPS to the *holo*-form is catalyzed by PPtases under the consumption of CoA. Subsequently, amino acid substrates are specifically recognized and activated by the corresponding A-domains as aminoacyl adenylates. This reactive species is then transferred and immobilized on the PCP-bound ppan cofactor. C-domain-mediated condensation of the template bound substrates affords a dipeptidyl intermediate which can be furthermore elongated and released from the synthetase, ultimately leading to the assembled NRP.

### 1.1.1 The Essential Domains

Domains required for the elongation of the oligopeptide during nonribosomal assembly are considered essential. These domains are the adenylation (A)-domain, the condensation (C)-domain and the peptidyl-carrier-protein (PCP). A minimal module consisting of one of each essential domain is capable of elongating the oligopeptide by one unit. The following chapter gives a more concise description of each domain.

#### 1.1.1.1 The Adenylation (A)-Domain

The adenylation domain is responsible for substrate recognition and activation. The high degree of selectivity of this class of enzymes defines the primary sequence of the readily assembled oligopeptide.[32] After specific recognition of an amino acid or carboxylic acid, the A-domain catalyzes the formation of a reactive aminoacyl adenylate in the presence of $Mg^{2+}$ and ATP.[33] This reaction is analogous to the activation of amino acids by aminoacyl-tRNA synthetases during the ribosomal assembly of peptides. Although both enzymes catalyze the same reaction they do not exhibit any structural or sequential homology.[34]

A-domains represent autonomous catalytic units that maintain their activity when separated from the enzymatic template and display a higher degree of substrate tolerance compared to aminoacyl-tRNA synthetases due to the lack of proof-reading-mechanisms.[35] This relaxed substrate specificity can result in the assembly of differing NRPs by the same NRPS machinery. A-domains usually consist of ~550 aa and can be subdivided into the *N*-terminal core domain ($A_{core}$ ~450 aa) and a smaller *C*-terminal subdomain ($A_{sub}$ ~100 aa). Structural analysis of the PheA A-domain, dissected from the gramicidin S-synthetase GrsA soaked with the cognate substrate and ATP, gave insights into the relative position of the active-site and the residues responsible for the recognition of the substrate.[36] In combination with biochemical and bioinformatic analysis of a set of A-domains, the specificity-conferring code of A-domains was established. The ten amino acids constituting this code are responsible for substrate binding within the active-site. Application of this code for the bioinformatic analysis of A-domains allows the prediction of the putative substrates activated by the corresponding A-domains. This approach is currently applied in the prediction and isolation of NRPS-derived products via genome mining methods.[37-39]

#### 1.1.1.2 The Peptidyl-Carrier-Protein (PCP)

The peptidyl-carrier-protein (PCP) represents the universal transporter unit within modular NRPSs and is usually located *C*-terminally of A-domains. It is responsible for the covalent tethering of monomeric building blocks or peptidyl intermediates and translocation of the growing NRP chain.[40] The substrates or intermediates are covalently immobilized on the sulfhydryl-group of a phosphopantetheine cofactor (ppan) which is attached to the highly conserved serine residue of the core-T motif (GGxS). Conversion of the inactive *apo*-PCP to the *holo*-PCP results from a posttranslational modification of the PCP, mediated by phosphopantetheinyl-transferases (PPtases).[29,41] PPtases catalyze the nucleophilic attack of the hydroxyl-group of a serine residue onto the β-phosphate group of coenzyme A, releasing 3',5'-adenosinediphosphate (3',5'-ADP). PCPs share a high degree of sequential and structural homology

to acyl-carrier-proteins (ACP) found in fatty acid synthases (FASs) and polyketide synthases (PKSs) and are designated aryl-carrier-proteins (ArCP) when the cognate substrates are aryl acids as found in the biosynthesis of triostin A.[42] NMR structural studies of $TycC_3$-PCP revealed the PCP to adopt three different and slowly converting conformations: the *apo* (*A*), the *A*/*H* and the *holo* (*H*) state.[43] When the PCP is in the *apo*-state, both the *A* and the *A*/*H* state coexist whereas in the case of *holo*-PCP the *H* and the *A*/*H* state coexist. These dynamic conformational changes enable the interaction of the ppan-bound substrates and intermediates with neighbouring domains. An additional finding of this study was that the ppan cofactor of *holo*-PCP and especially the terminal sulfhydryl-group is able to move approximately 16 Å, which confirmed the postulated swinging mode of the ppan prosthetic group.

#### 1.1.1.3 The Condensation (C)-Domain

The condensation (C)-domain, which comprises ~450 aa, is an additional essential domain within the multimodular NRPS template.[44] C-domains catalyze peptide bond formation between an electrophilic PCP-bound donor substrate presented from the *N*-terminal side of the synthetase and a nucleophilic PCP-bound acceptor substrate. The formation of the peptide bond is initiated by the nucleophilic attack of the α-amino group of the acceptor substrate onto the peptidyl-*S*-PCP thioester of the donor substrate.[45-47] The elongated peptide is subsequently located on the downstream PCP-domain and can be subjected to further elongation- and translocalization-steps. All condensation domains have been found to operate unidirectional by translocalizing the elongation product towards the *C*-terminal termination module.The number of C-domains located within the synthetase usually correlates with the number of peptide bonds found in the readily assembled product. This model implies that NRPS-based assembly follows a linear logic. In the case of nonlinearly operation NRPSs, exemplified by CchH involved in coelichelin biosynthesis, C-domains have to exhibit other mechanisms of inter-domain communication.[38] Structural and bioinformatic analysis of C-domains identified a highly conserved catalytic His-residue located in the C3-core motif (MHHxxxDG(WV)S).[48] Mutational studies of the $TycC_6$ PCP-C didomain suggest that the second histidine of the HHxxxDG motif may catalyze the deprotonation of the α-ammonium group of the acceptor substrate to enhance the nucleophilicity for the nucleophilic attack onto the donor substrate thioester, but the exact mechanism remains to be elucidated.[49] C-domains are functionally diverse catalysts that are involved in various biosynthetic processes. In the case of arthrofactin biosynthesis, a dual condensation/epimerization-domain could be identified which epimerizes the upstream aminoacyl/peptidyl moiety before condensation when the condensation domain is simultaneously presented with the L-aminoacyl-*S*-pantetheinyl acceptor.[50] The C-domain

located at the *N*-terminus of the SrfA-A synthetase, generally designated $C^{III}$-domain, was recently shown to be involved in the initiation of surfactin biosynthesis.(Femke Kraas, Philipps-University Marburg, unpublished results).[51]

#### 1.1.1.4 Termination of Nonribosomal Peptide Assembly

Termination of NRP biosynthesis is achieved by the catalytic or autocatalytic cleavage of the assembled oligopeptide from the enzymatic template. A variety of release mechanisms has evolved, contributing to the structural diversity of NRPs. The most commonly found release strategy in NRP biosynthesis is mediated by thioesterase (TE)-domains located at the *C*-terminus of the synthetase. This class of enzymes displays a high degree of sequential and structural homology to $\alpha/\beta$-hydrolases.[52] Depending on the nature of the nucleophile that is employed in the TE-mediated release, linear, cyclic or branched cyclic peptides result. TE-domains can catalyze the macrocyclization of NRPs through the intramolecular nucleophilic attack of a side-chain nucleophile onto the acyl-O-TE-oxoester intermediate.[53] As this macrocyclization strategy is a major part of the work presented herein, it is discussed in detail in chapter 1.5. The hydrolytic release of the template-bound NRP represents an alternative release strategy, catalyzed by hydrolyzing TE-domains. This hydrolytic release results in a linear peptide and has been observed during the biosynthesis of vancomycin, complestatin or yersiniabactin.[54-56] In addition, *in trans* acting TE-domains were postulated to be involved in mannopeptimycin and coelichelin biosynthesis.[38,57] The reduction of the *C*-terminal carboxyl group to afford the corresponding aldehyde or aminoalcohol represents an additional release strategy that has been observed during nostocyclopeptide and linear gramicidin biosynthesis.[58-59] This reduction is catalyzed by *C*-terminally located reductase (R)-domains in a NAD(P)H-dependent manner. In the case of the macrocyclic nostocyclopeptide, reductive release of the linear peptide aldehyde triggers macrocyclization via an intramolecular imination.[59] In several NRPSs, the *C*-terminal TE- or R-domain is replaced by a condensation domain that is postulated to mediate the cyclorelease of the PCP-bound peptide. These *C*-terminal C-domains were identified in the biosynthetic gene clusters responsible for the assembly of cyclosporine or the DKP-containing toxins thaxtomin and fumitremorgin.[60-62]

### 1.2 *In Cis* Operating Optional Domains

In addition to the essential core domains, further domains could be identified and biochemically characterized which are responsible for the *in cis* modification of the peptidic backbone and contribute to the structural and functional diversity of NRPS-derived natural products.

#### 1.2.1 The Epimerization (E)-Domain

A key feature of NRPs is the incorporation of D-configured amino acids in the peptidic backbone. The presence of D-configured amino acids can be observed in numerous NRPS-derived compounds, *e.g.* vancomycin or daptomycin.[54,63] D-isomers of amino acids not only affect the bioactivity of the compound but also prevent the proteolytic degradation of the peptide.[23] The conversion of the L-amino acid to the D-isomer is generally mediated by epimerization (E)-domains, comprising ~450 aa, located downstream of adjacent PCPs. E-domains epimerize the L-configured substrate after immobilization as an aminoacyl-*S*-PCP intermediate. E-domains located in initiation modules generate a mixture of PCP-*S*-L,D-monomers with the downstream C-domain selecting the D-isomer for the initiation of peptide assembly. E-domains embedded in elongation modules were shown to catalyze the epimerization of the peptidyl-*S*-PCP intermediate prior to translocalization.[64] C-domains located downstream of such E-domains are also specific for the D-configured substrate and assure a selective incorporation of the D-isomer into the product.[44] Bioinformatic analysis of E-domains revealed the presence of a catalytically active His-residue located in the E3-motif (HHxxxDG) which is believed to mediate base-catalyzed $C_\alpha$ epimerization.[51]

#### 1.2.2 The Methylation (MT)-Domain

Methylation (MT)-domains, also termed methyltransferases, catalyze the transfer of a methyl group from the methyl group donor *S*-adenosylmethionine to carbon, nitrogen or oxygen atoms of NRPs *in cis*. MT-domains can be distinguished by the site of methylation and are designated as C-MT, N-MT or O-MT, respectively.[65] Methylation can be observed in a set of NRPS-derived products such as pristinamycin, actinomycin or thiocoraline.[66-68] MT-domains comprise ~420 aa and have a didomain structure with the *N*-terminal domain containing the SAM-binding site and the *C*-terminal domain being responsible for substrate binding. MT-domains are usually inserted into the *C*-terminal region of A-domains and contain at least three highly conserved motifs.[65] *N*-methylation of the aminoacyl-*S*-PCP intermediate occurs prior to condensation as observed during the characterization of actinomycin or cyclosporin biosynthesis.[67,69]

### 1.2.3 The Formylation (F)-Domain

*N*-formylation of NRPs has only been observed during the biosynthesis of linear gramicidin by LgrA *(Bacillus brevis* ATCC 8185) and during anabaenopeptilide assembly by ApdA (*Anabaena* str. 90).[70-71] Formylation of the *N*-terminal α-amino group is catalyzed by the formylation domain, which is located at the *N*-terminus of the corresponding synthetase and comprises ~200 aa. The formylation domain catalyzes formyl group transfer from formyltetrahydrofolate ($fH_4F$) onto the α-amino functionality of the amino acid using both cofactors ($N^{10}$- and $N^5$-$fH_4F$). The *in vitro* characterization of the LgrA F-domain revealed the enzyme to catalyze α-*N*-formylation of PCP-bound L-Val or L-Ile. In addition, the necessity of the formylated starter unit for the initiation of gramicidin biosynthesis was observed.[70]

### 1.2.4 The Cyclization (Cy)-Domain

Heterocyclization of cysteine, serine or threonine residues to the corresponding five-membered thiazoline, oxazoline or methyloxazoline heterocycles is a structural feature found in several NRPs. The introduction of these structures increases the diversity of the natural product and furthermore rigidifies the peptide backbone. The enzymes mediating this heterocyclization are modified C-domains that catalyze peptide bond formation and a subsequent nucleophilic attack of hydroxyl- or sulfhydryl-side chains onto the peptide bond.[72] The dehydratation of the resulting intermediary structure gives rise to the corresponding heterocycle.[73] Heterocycles can be found in yersiniabactin, bleomycin and bacitracin and are often essential for the chelation of iron.[73-75] Cy-domains can be associated with oxidation (Ox)-domains, catalyzing the FMN-dependent two-electron-oxidation of dihydroaromatic thiazolines or oxazolines to thermodynamically more stable thiazoles or oxazoles.[72,76] Reduction of thiazolines or oxazolines to afford tetrahydrated thiazolidines or oxazolidines is mediated by *in trans* operating reductases. These NAD(P)H-dependent enzymes recognize and directly reduce the PCP-bound substrate.[77]

## 1.3 Tailoring and Modification of NRPS Building Blocks

The structural diversity of NRPS-derived products arises not only from modifications *in cis*, but also from modifying enzymes generating unusual building blocks or mediating postsynthetic modifications of the natural product (Figure 1.4).[78-79] Unusual building blocks often represent tailored proteinogenic amino acids that are recognized and activated by the corresponding A-domains.[80] An alternative mechanism arises from the online-modification of the PCP-bound building block by *in trans* operating tailoring

enzymes.[30] Postsynthetic modifications can result from decoration with sugar moieties by glycosyltransferases (Gtfs), which in the case of the anti-tumor agent bleomycin contribute to the DNA-binding specificity (Figure 1.4). The following chapter gives a concise overview of common modifications found in NRPs and focuses on the enzymes mediating these modifications.

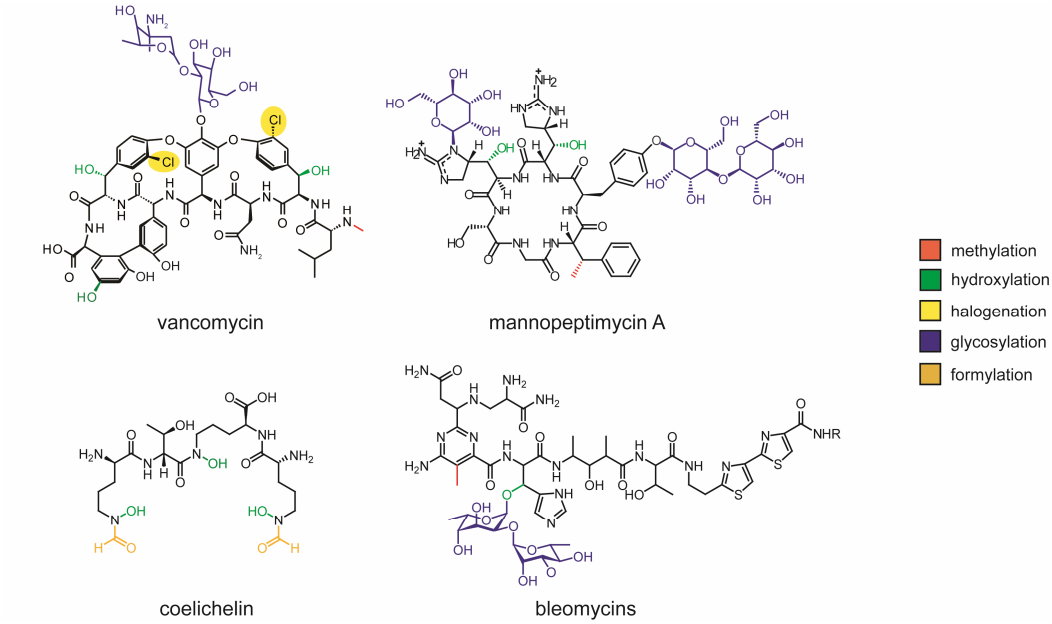



**Figure 1.4:** Examples for the structural diversity of peptides of nonribosomal origin. Modifications of the peptidic backbone include methylations, hydroxylations, halogenations, glycosylations and formylations. The peptides shown are vancomycin, mannopeptimycin A, coelichelin and the bleomycins. Modifications are accentuated in different colors.[38,79,81-82]

### 1.3.1 Methylation

Methylations are generally governed by *S*-adenosylmethionine (SAM)-dependent methyltransferases.[83] In addition to SAM, methyltransferases can also employ the cosubstrate methyl-tetrahydrofolate as the cognate methyl group donor.[84] Methyltransferases catalyze the methylation of carbon, nitrogen, oxygen or sulphur-atoms, converting SAM to *S*-adenosylhomocysteine (AdoHCy). In addition to *in cis* acting methyltransferases, a set of *in trans* acting SAM-dependent methyltransferases could be identified and characterized. These methyltransferases were shown to be involved in the biosynthesis of unnatural building blocks required for the assembly of NRPs. One of the first methyltransferases to be biochemically characterized *in vitro* was GlmT, encoded in the CDA biosynthetic gene cluster (*Streptomyces coelicolor* A3(2)) (Figure 1.5 A).[85-86] Intriguingly, GlmT did not catalyze the methylation of Glu directly. Instead, the substrate for GlmT-mediated methyltransfer was shown to be α-ketoglutarate (α-KG), leading to 3-methyl-2-oxoglutarate. Subsequently, an additional transamination step is required

for the conversion of 3-methyl-2-oxoglutarate to 3-methyl-L-glutamate (MeGlu). A branched-chain aminotransferase from the primary metabolism of *S. coelicolor* A3(2) was demonstrated to catalyze this reaction with L-Val being the amino group donor. Recently, the methyltransferase MppJ, encoded in the mannopeptimycin (*Streptomyces hygroscopicus*) gene cluster, was characterized *in vitro*.[82,87] (Figure 1.5 B) It was shown that the cognate substrate for $C_\beta$-methylation is phenylpyruvate instead of L-Phe. The methylation mechanism is analogous to GlmT-mediated methylation of α-ketoglutarate. Interestingly, the methylation and subsequent transamination of phenylpyruvate generated two diastereomers, (2*S*,3*R*)- and (2*S*,3*S*)-3-methyl-phenylalanine although solely the latter is found in mannopeptimycin.[82]

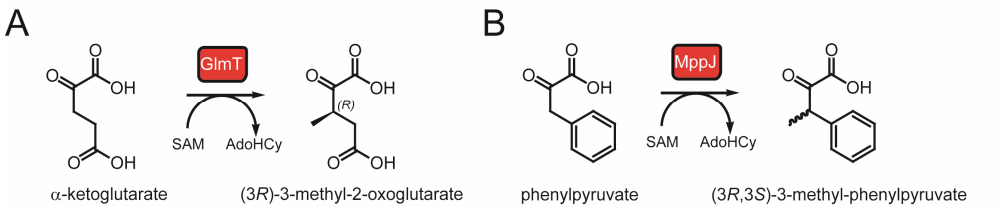



**Figure 1.5:** Examples for the modification of free precursors by SAM-dependent methyltransferases. The methyl group donor SAM is converted to the corresponding AdoHCy. A) GlmT-mediated methylation of α-KG as a key step of the biosynthetic route for the generation of the 3-methyl-L-glutamate residue found in CDA. B) MppJ-mediated methylation of phenylpyruvate during mannopeptimycin biosynthesis.

### 1.3.2 Halogenation

Halogenases catalyze the functionalization of carbon atoms and can be subdivided into four classes: $FADH_2$- and nonheme Fe(II)/α-KG-dependent halogenases employ molecular oxygen for the oxidation of the halogen, whereas vanadium- and heme-iron-dependent halogenases employ hydrogenperoxide for the generation of the reactive halogen species.[88] Common halogenations are chlorination, bromination or iodination, but fluorination has also been observed in the biosynthesis of 5'-fluoro-5'-deoxyfluoroadenosine (5'-FDA).[89] Examples for halogenated NRPS-derived natural products are chondramide, syringomycin and the kutznerides. In the case of the highly cyctotoxic chondramide (*Chondromyces crocatus* Cm c5), the $FADH_2$-dependent halogenase CmdE catalyzes chlorination of free L-Trp to generate 2-chloro-L-Trp (Figure 1.6 A).[90] An alternative halogenation mechanism is employed during the biosynthesis of syringomycin (*Pseudomonas syringaea*). The nonheme Fe(II)/α-KG-dependent halogenase SyrB2 catalyzes the chlorination of PCP-bound L-Thr, which is subsequently participating in peptide elongation steps (Figure 1.6 C).[91] Halogenation of a PCP-bound L-Ile has also been shown to be essential for the biosynthesis of (1*S*,2*R*)-allocoronamic acid, an unnatural monomer found in the

kutznerides (*Kutzneria* sp. 744).[92-93] The nonheme Fe(II)-dependent halogenase KtzD was demonstrated to catalyze the formation of a PCP-bound γ-chloroisoleucyl intermediate, which is cyclized to the final product by KtzA (Figure 1.6 B).[93]

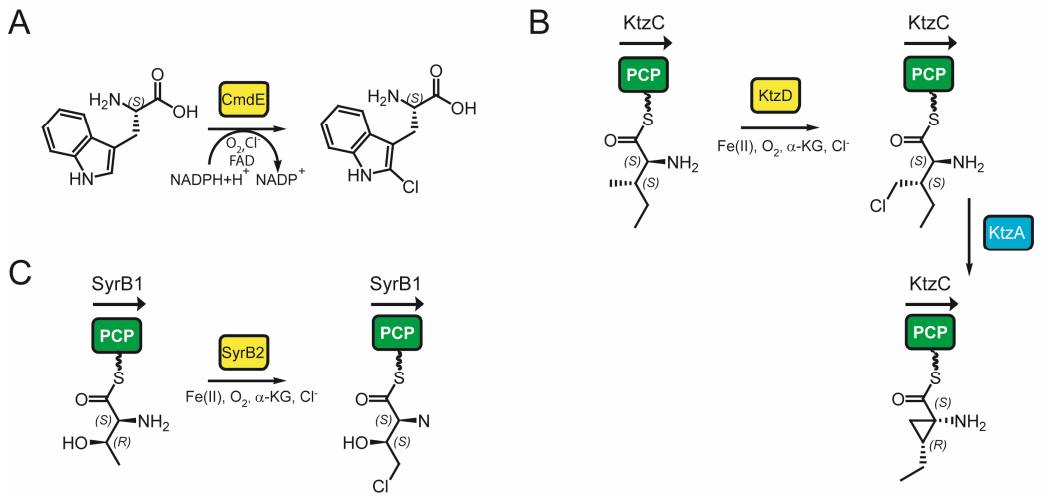



**Figure 1.6:** Examples for the modification of free precursors or PCP-bound substrates by $FADH_2$- and nonheme Fe(II)/α-KG-dependent halogenase during the biosynthesis of NRPs. A) The FAD-dependent halogenase CmdE mediates chlorination of free L-Trp during chondramide biosynthesis. B) The nonheme Fe(II)/α-KG-dependent halogenase KtzD governs γ-chlorination of PCP-bound L-Ile *in trans*. In a second reaction the acyl-CoA dehydrogenase-like flavoprotein KtzA catalyzes cyclization to afford the final product. C) SyrB2-mediated chlorination of PCP-bound L-Thr during syringomycin biosynthesis.

### 1.3.3 Hydroxylation

Hydroxylation reactions are generally mediated by three enzyme classes: FAD-dependent monooxygenases, nonheme Fe(II)-oxygenases and heme Fe(II)-oxygenases. The class of FAD-dependent monooxygenases catalyzes a broad variety of oxygenations, including epoxidations.[94-95] These enzymes require the reduced FAD-cofactor for catalysis with the electrons for cofactor reduction usually supplied by NADH or NADPH. This class of FAD-dependent monooxygenases is designated external flavoprotein monooxygenases. The enzymes are termed internal flavoprotein monooxygenases if cofactor reduction is mediated by the cognate substrate.[94] Alike the FAD-dependent monooxygenases the nonheme Fe(II)-oxygenases catalyze a broad spectrum of oxygenations. Nonheme Fe(II)/α-KG-dependent monooxygenases represent the largest subgroup of the aforementioned oxygenases and couple the oxidative conversion of the substrate with the decarboxylation of the cosubstrate α-ketoglutarate to succinate and carbon dioxide.[96-97]

*In trans* hydroxylation reactions for the generation of unnatural monomers during nonribosomal assembly of natural products have been observed for FAD-dependent monooxygenases and nonheme Fe(II)/αKG-dependent monooxygenases. FAD-dependent monooxygenases are often involved in the biosynthesis of hydroxamate-type siderophores. The hydroxylase CchB was shown to catalyze δ-*N*-hydroxylation of L-ornithine for the generation of δ-*N*-hydroxy-L-ornithine (hOrn) during coelichelin biosynthesis (*Streptomyces coelicolor* A3(2)) and IucD was demonstrated to be involved in the generation of ε-*N*-hydroxy-L-lysine (hLys) during aerobactin biosynthesis (*E. coli* K-12) (Figure 1.7 A/B).[98-99]

Nonheme Fe(II)/α-KG-dependent monooxygenases involved in the $C_\beta$-hydroxylation of either free amino acids or PCP-bound substrates were identified in the biosynthetic pathways of CDA, viomycin and the kutznerides. The monooxygenase AsnO catalyzes the regio- and stereospecific hydroxylation of free L-Asn during CDA biosynthesis, whereas VioC catalyzes the specific $C_\beta$-hydroxylation of L-arginine in viomycin biosynthesis (*Streptomyces* sp. strain ATCC 11861) (Figure 1.7 D/E).[100-101] In constrast, KtzO and KtzP, involved in the biosynthesis of the kutznerides, only modify PCP-bound L-Glu substrates. KtzO mediates the generation of the *threo* diastereomer of β-hydroxy-glutamate, whereas KtzP catalyzes the generation of the *erythro* diastereomer.[102]

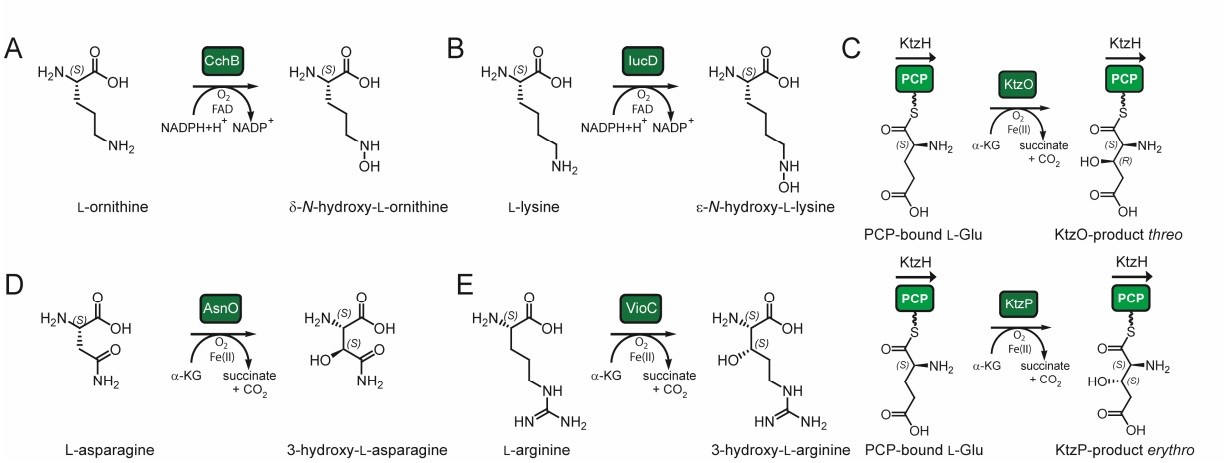



**Figure 1.7:** Examples for the modification of free building blocks or PCP-bound substrates by FAD- and nonheme Fe(II)/α-KG-dependent monooxygenases during the biosynthesis of NRPs. A) The FAD-dependent monooxygenase CchB mediates δ-*N*-hydroxylation of L-Orn during coelichelin biosynthesis. B) IucD catalyzes the hydroxylation of L-Lys during aerobactin assembly. C) Hydroxylation of PCP-bound L-Glu by the nonheme Fe(II)/α-KG-dependent monooxygenases KtzO and KtzP *in trans*. D) $C_\beta$-hydroxylation of L-Asn by AsnO. E) Viomycin biosynthesis requires the hydroxylation of L-Arg by VioC.

### 1.4 Classification of Nonribosomal Assembly Line Logic

NRPS assembly lines can be subdivided into three groups based on the specific use of domain order and the multimodular template which are discussed in detail in the following chapter.

#### 1.4.1 Linear NRPS-Assembly Line Logic (Type A)

The most extensively characterized NRPS-system is the linearly operating assembly line (type A). The common domain and module organization corresponds to the minimal module (C-A-PCP), which is responsible for the incorporation of one building block into the growing peptide chain. The order of elongation modules and the number of adenylation-domains directly correlates with the primary sequence of the assembled oligopeptide. Product release is commonly catalyzed by a *C*-terminal TE-domain which catalyzes hydrolytic release or macrocyclization. Examples for NRPs derived from type A NRPS assembly lines are surfactin, tyrocidine, cyclomarazines or daptomycin.[36,103-105] Assembly of daptomycin follows a linear logic, involves *in trans* operating modifying enzymes for the generation of unnatural building blocks and serves as an example for the linear biosynthetic logic (Figure 1.8).[86,106]

#### 1.4.2 Iterative NRPS-Assembly Line Logic (Type B)

In contrast to type A NRPSs, the iteratively operating NRPS assembly lines (type B) use their enzymatic template more than once for the biosynthesis of the corresponding product. The module and domain organization is analogous to type A NRPS-systems. A key feature of type B assembly lines is the termination module which is responsible for the covalent connection of the repetitive units and subsequent product release. Examples for iteratively assembled NRPs are the siderophore enterobactin, which is a cyclic trimer consisting of 2,3-DHB-seryl-subunits, the macrocyclic decapeptide gramicidin, the macrolactone enniatin or the chromodepsipeptide triostin A.[58,107-109] The mechanism of iterative NRP biosynthesis is discussed in detail (Chapter 1.5).

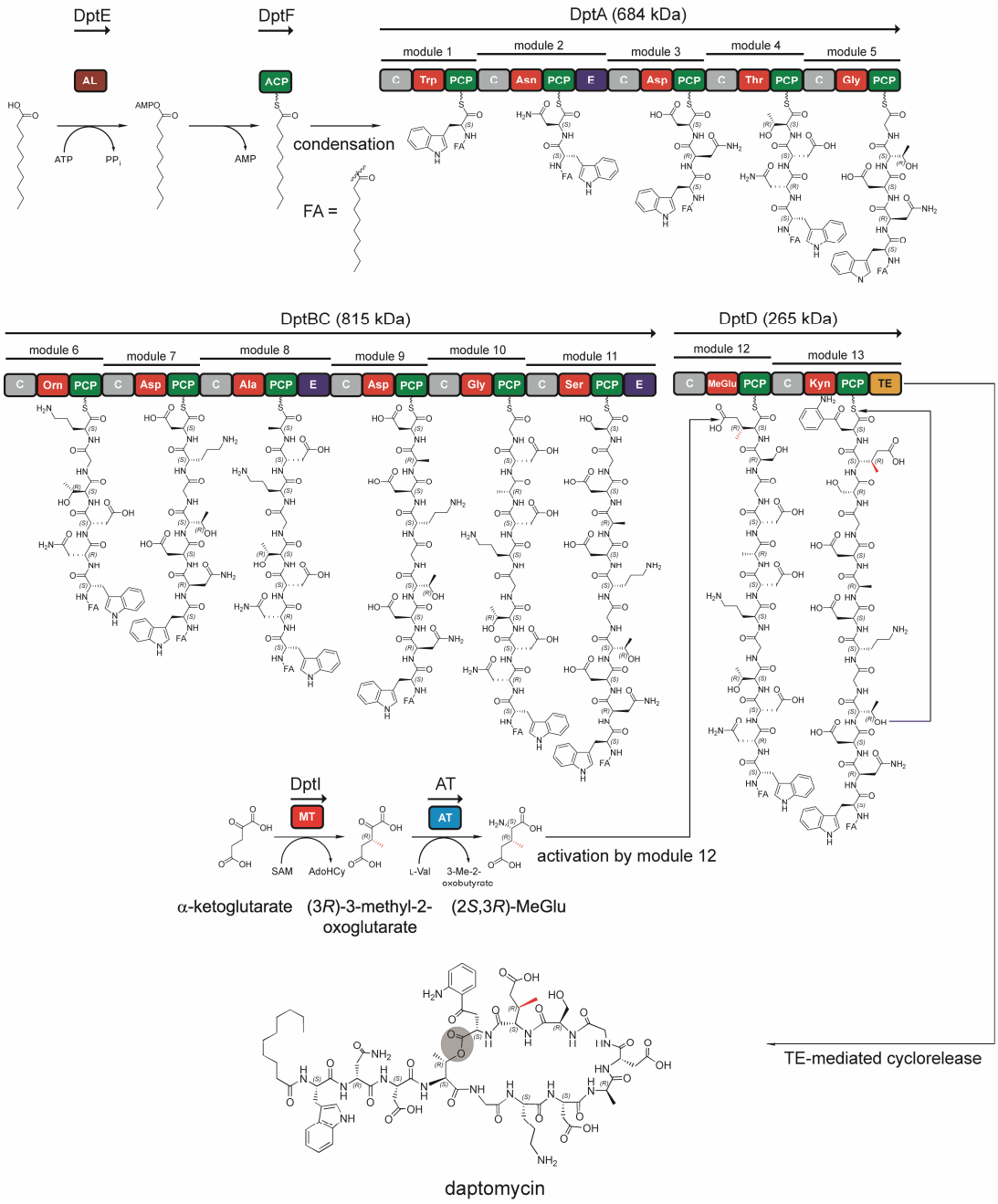



**Figure 1.8:** The assembly of daptomycin is governed by the three linearly operating nonribosomal peptide synthetases DptA, DptBC and DptD, comprising 43 catalytic domains. Initiation of daptomycin biosynthesis is mediated by DptE (acyl-CoA ligase) and DptF (ACP), both responsible for the activation and incorporation of the fatty acid (FA) moiety. In this model the *N*-terminal $C^{III}$-domain of DptA catalyzes the transfer of the DptF-bound FA onto the α-amino group of $Trp_1$. The synthesis of the $MeGlu_{12}$ residue is carried out by the SAM-dependent methyltransferase DptI and a currently unknown aminotransferase (AT). Cyclorelease is mediated by the *C*-terminal thioesterase domain of DptD. The cyclization position is accentuated in grey.

### 1.4.3 Nonlinear NRPS-Assembly Line Logic (Type C)

Nonlinear NRPSs (type C) differ in their organization of modules and domains from type A and type B NRPSs. Domain organization within a module does not follow a (C-A-PCP) logic and the number and organization of modules or domains does not reflect the primary sequence of the assembled product. In addition, lone-standing individual domains are often involved in product assembly. The prediction of secondary metabolite structure is often impeded by the nonlinearity of type C assembly lines. A prominent example of a nonlinear type C NRPS-system is the Cgc assembly line responsible for the production of the pyrrole-amide antibiotic congocidin (*Streptomyces ambofaciens*). Within the operon only one modular NRPS, Cgc18 (A-PCP-C), is encoded but congocidin consists of four building blocks. Additional condensation steps are putatively carried out by lone-standing C-domains (Figure 1.9).[110]

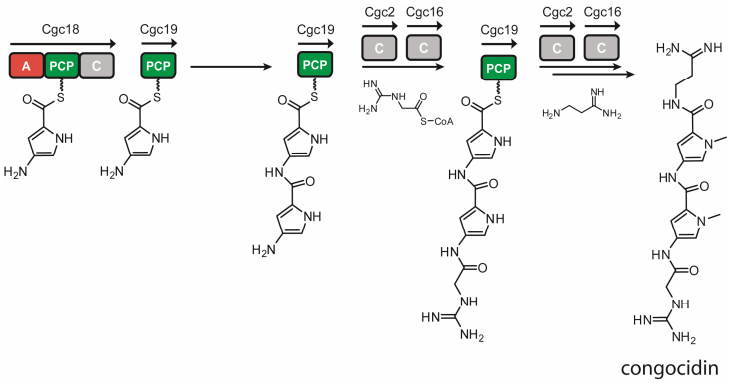



**Figure 1.9:** The postulated mechanism for the nonlinear assembly of the pyrrole-amide antibiotic congocidin. Peptide bond formation is mediated by the modular NRPS Cgc18 and the lone-standing C-domains Cgc2 and Cgc16.

### 1.5 Iterative Termination Strategies During Nonribosomal Peptide Assembly

The key step in iterative NRP assembly is the termination reaction in which repetitive units are covalently linked and released from the enzymatic template. This cyclodimerization or cyclotrimerization is usually governed by iteratively operating thioesterase (TE)-domains in bacterial systems.[111-112] In contrast, fungal systems often employ *C*-terminal C-domains instead of TE-domains for the catalysis of the cyclization reaction. As a major part of the work presented herein focuses on the biochemical characterization of an iteratively acting TE-domain derived from the thiocoraline biosynthetic pathway, the following chapters give a concise overview of mechanistic and structural aspects of TE-domains.

### 1.5.1 The Thioesterase – Structural and Mechanistic Aspects

TE-domains found in NRPSs are structural representatives of the α/β-hydrolase superfamily which share a common fold, but exhibit a variety of activities.[113] The common fold consists of a core of eight mainly parallel oriented β-sheets, which are flanked by one α-helical segment each on the *N*- or *C*-terminal side (Figure 1.10 A).

NRPS-derived TE-domains are characterized by an insertion of three α-helices which together constitute a lid-region putatively involved in substrate recognition.[114] Based on the three- dimensional structure of the TE-domain involved in surfactin biosynthesis, it was postulated that the lid-region experiences a conformational change upon binding of the substrate, resulting in either an opened- or closed-state.[114] In contrast, the TE-domain involved in fengycin biosynthesis contains a canyon-like active site.[115]

The tertiary structure of NRPS TE-domains induces the formation of an active site, in which the catalytically active serine of the GxSxG core motif is located (Figure 1.10 B). The formation of a catalytic triad consisting of Ser-His-Asp(Glu) leads to the deprotonation of the active-site serine and generates a highly reactive oxyanion species. TE-mediated cyclization is initiated by the nucleophilic attack of the serine oxyanion onto the adjacent PCP-bound peptidyl thioester leading to the conversion of the peptidyl thioester to an acyl-O-TE oxoester intermediate.[116] The coordination of this peptidyl oxoester results in the spatial proximity of an internal nucleophile and the carbonyl group of the acyl-O-TE. A subsequent intramolecular nucleophilic attack of the nucleophile gives rise to a negative tetrahedral transition state. The tetrahedral transition state is stabilized by interaction of the oxyanion with the amide backbone of adjacent amino acids. This structural component is designated oxyanion hole and was initially postulated for the serine protease class of α/β-hydrolases.[117]

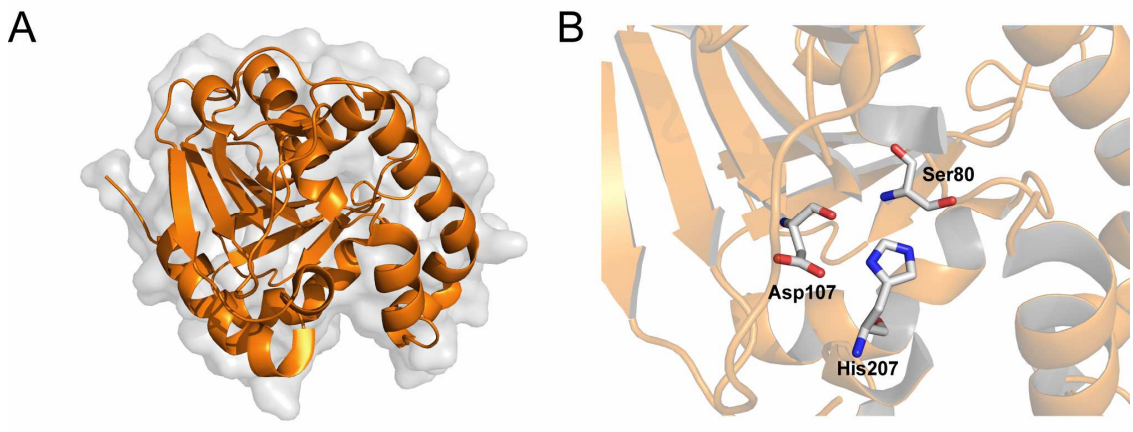



**Figure 1.10:** The three-dimensional structure of the SrfA-C TE-domain (PDB code: 1JMK). A) The α/β-hydrolase fold is characterized by eight mainly parallel oriented β-sheets, which are flanked by one α-helical segment each on the *N*- or *C*-terminal side. B) The catalytic triad consisting of Ser80, His207 and Asp107.

NRPS thioesterases (cyclases) have been shown to catalyze macrolactamization and macrolactonization affording both cyclic and branched cyclic products.[118-119] Macrolactonization requires a hydroxyl group provided either by Ser, Thr or Tyr residues or by hydroxylated fatty acid residues.[53,116,120] In contrast, macrolactamization requires amino groups provided by the α-amino group of the *N*-terminal amino acid, Lys, Orn or diaminopropionate (Dap) residues.[121-122] An unprecedented macrocyclization variant is macrothiolactonization resulting in thioester-linked natural products. Only two NRPS-derived natural products with thioester linkage have been identified: the anti-tumor compound thiocoraline and BE-22179, both members of the macrocyclic chromodepsipeptides.[68,123] The intrinsic pseudosymmetry of these natural products arises from a repetitive head-to-tail macrocyclization putatively mediated by iteratively operating thioesterases which is addressed in the following chapter.

### 1.5.2 Iterative Termination Strategies

The number of modules defining an iteratively operating NRPS-system does not correlate with the primary sequence of the assembled product. Biosynthesis of these natural products therefore requires the iterative use of the corresponding template and dimerization or trimerization of the individual units. The general catalyst for this combined ligation and cyclization reaction was identified to be the *C*-terminal TE-domain located in the termination module of the NRPS.

This iterative termination strategy has been observed and characterized in enterobactin biosynthesis (*E. coli*) and in gramicidin S biosynthesis (*Bacillus brevis*).[111-112] The first results that confirmed the iterative assembly of a NRP were obtained during the mass spectrometry based analysis of enterobactin

biosynthesis (Figure 1.11). Enterobactin is a trilactone siderophore that consists of three identical 2,3-DHB-seryl units. The macrocyclic structure arises from three ester bonds between the side-chain hydroxyl group of serine and the carboxyl group of a second monomer.[107] Generation of an EntF TE-mutant in which the His-residue participating in the catalytic triad was substituted by Ala led to the deceleration of enterobactin biosynthesis.[111] MS-analysis revealed the presence of TE-bound monomeric and dimeric units. It was postulated that the synthetase presents a monomer to the TE-domain which is subsequently transferred onto the active-site serine. The monomer remains TE-bound until a second monomer is presented on the adjacent PCP. TE-mediated ligation affords the TE-bound dimer and a third ligation reaction gives rise to the TE-bound trimer which is cyclized and released from the enzymatic template.

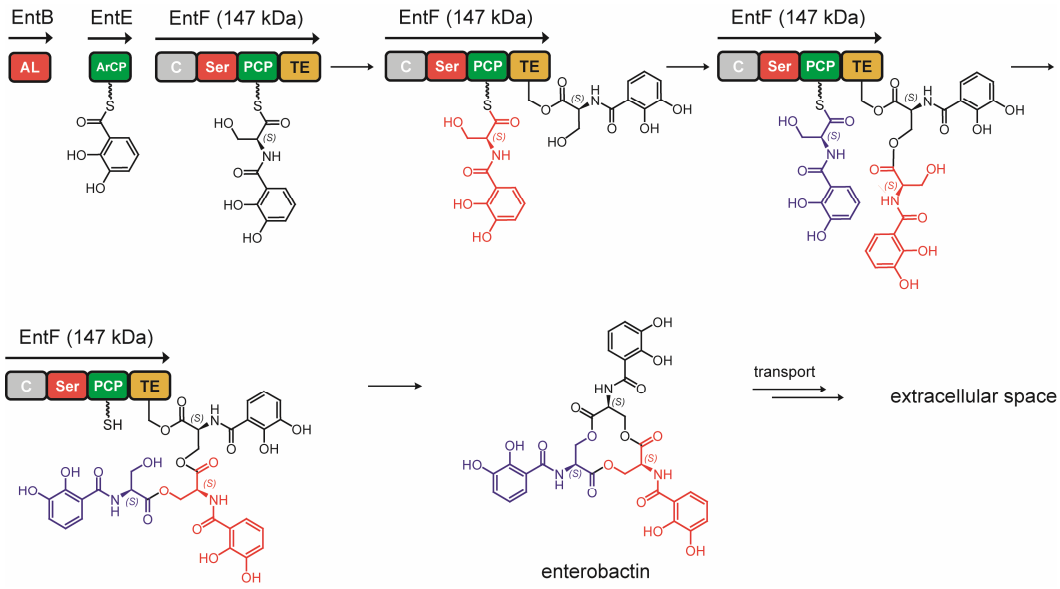



**Figure 1.11:** The postulated mechanism for enterobactin assembly. The lone-standing AMP-ligase EntB activates 2,3-DHB which is transferred to the synthetase EntF by the ArCP EntE. Assembly of the trilactone enterobactin is realized by iterative dimerization and trimerization of the monomeric units. Macrocyclization ultimately leads to the formation of enterobactin. The monomeric units are accentuated in black, red and blue.

The *in vitro*-characterization of the TE-domain involved in the assembly of the macrolactam decapeptide gramicidin S granted further insights into the iterative mechanism.[112] To analyze the combined ligation and cyclization reaction a set of pentapeptidyl-substrates, resembling cognate monomeric units, was synthesized via solid-phase peptide synthesis (SPPS) and *C*-terminally activated as thioesters (Chapter 1.7.1). The *C*-terminal SNAC activation mimicked the naturally occurring ppan cofactor of the

PCP-bound peptidyl substrate. Incubation of the substrates with the recombinant GrsB TE gave rise to macrocyclic gramicidin S derivatives. The detection of a *C*-terminally activated linear decapeptide led to the postulation of two unique elongation mechanisms. The forward mechanism is characterized by the nucleophilic attack of a TE-bound pentapeptidyl intermediate onto the thioester of the PCP-bound pentapeptide. The resulting TE-bound linear decapeptide is subsequently cyclized and released from the synthetase. In contrast, the backward mechanism comprises an additional step. In the first reaction the α-amino group of the PCP-tethered pentapeptide attacks the acyl-O-TE oxoester of the TE-bound pentapeptide, resulting in a PCP-bound linear decapeptide (Figure 1.12). This intermediate is subsequently transferred onto the active-site serine and macrocyclized. The detection of the *C*-terminally activated decapeptide, resembling the PCP-bound decapeptide of the native system, confirms iteratively operating TE-domains to follow a backward mechanism logic.[112]

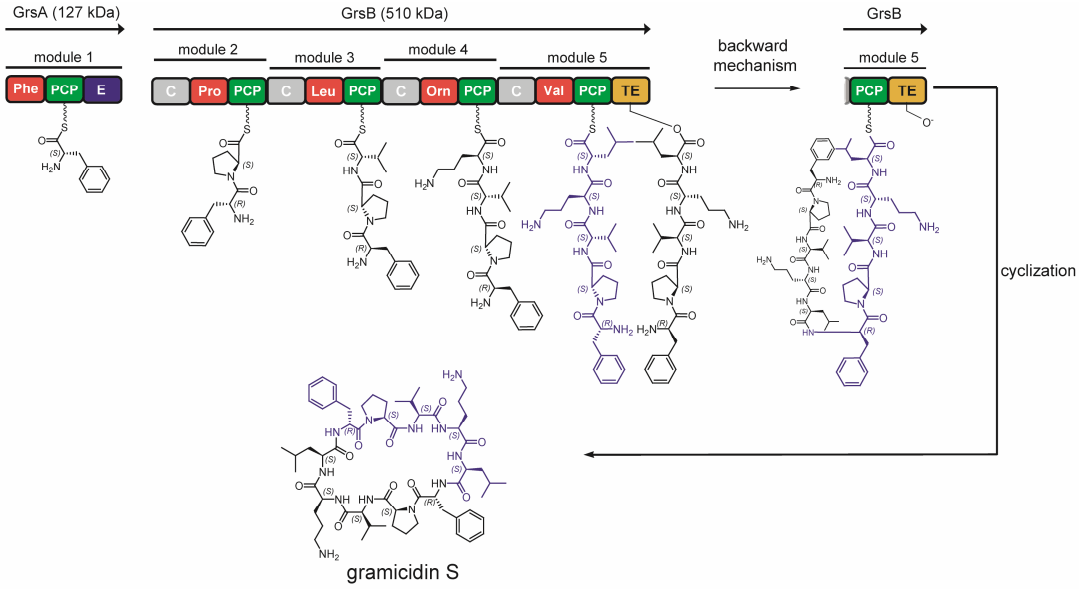



**Figure 1.12:** Iterative biosynthesis of the decapeptide gramicidin S by GrsA and GrsB. Termination of NRP assembly follows the backward mechanism and combines a ligation and cyclization reaction. The PCP-bound pentapeptide (accentuated in blue) conducts a nucleophilic attack onto the acyl-O-TE oxoester of the TE-bound pentapeptide resulting in a PCP-bound linear decapeptide. Subsequent TE-mediated cyclization affords the macrolactam gramicidin S.

### 1.6 Chromodepsipeptides and Siderophores – Peptides of Nonribosomal Origin

Nonribosomally assembled compounds represent a natural product class with a remarkable structural and pharmacological spectrum. As a result, several NRPs have been approved by the FDA as antibacterial, antifungal, antitumor or immunosuppressive agents (Table S1). The macrocyclic chromodepsipeptides are a class of NRPs that are extensively investigated due to their DNA-bisintercalative properties, which predestine them as antitumor agents.[124] The following chapter describes the structural features of chromodepsipeptides, their biosynthesis and the proposed mode of action of thiocoraline.

Bacteria produce and secrete small, iron-scavenging compounds in response to iron starvation. These compounds, designated siderophores, display a high affinity for the chelation of iron and are assembled in NRPS-dependent or NRPS-independent manners. Since siderophores often function as virulence factors in pathogens, the interest in the structural and functional characterization of these compounds is growing and may result in the synthesis of specific inhibitors based on the structure of the pathogen siderophore.[125] The following chapter gives insights into the structural and functional properties of siderophores and the biosynthetic mechanisms yielding these products.

#### 1.6.1 Thiocoraline – A Member of the Macrocyclic Quinoline- and Quinoxaline Chromodepsipeptides

Among the iteratively assembled nonribosomal peptides the class of chromodepsipeptides encompasses a broad variety of structurally and functionally diverse compounds (Figure 1.13). These peptides were shown to bind to duplex DNA through a mechanism known as bisintercalation mediated by the twin chromophores attached to the macrocyclic molecule.[124,126-128] Chromodepsipeptides share a common peptidic scaffold and a pseudosymmetrical structure resulting from the condensation of two symmetrical halves. This class can be subdivided into two main groups – the quinoxalines and the quinolines - depending on the chromophore moiety bound to the *N*-termini of each oligopeptide chain. Prominent members of the quinoxaline-group of chromodepsipeptides are echinomycin (antitumor) and triostin A (antitumor) which have been isolated from *Streptomyces lasaliensis* and *Streptomyces triostinicus*, respectively.[109,129] These compounds bind specifically to DNA via the insertion of the planar chromophore quinoxaline-2-carboxylic acid (QX), inhibiting transcription and replication that has led to the progression of echinomycin into clinical antitumor trials. Based on the intramolecular connectivity of these compounds, which results from either a disulfide bridge or a thioacetal linkage, these natural products can be furthermore classified as triostins (disulfide) or quinomycins (thioacetal). This rigidification of the macrocyclic core is contributing to the selectivity of the compound concerning DNA-binding as triostin A

and echinomycin share the same peptidic backbone but bind to different DNA-sequences.[130] The group of quinoline-chromodepsipeptides encompasses the natural products sandramycin (anti-HIV), luzopeptine A (anti-HIV), BE-22179 (antibiotic) and the recently discovered SW-163C (antitumor) isolated from *Nocardioides* sp. ATCC 39419, *Actinomadura luzonensis* nov. sp., *Streptomyces* sp. A22179 and *Streptomyces* sp. SNA15896.[123,131-133] The structural diversity among these compounds is furthermore increased by the incorporation of the unusual building blocks 4-hydroxy-2,3,4,5-tetrahydropiperazine-3-carboxylic acid (HTP), *N*-methyl-3-hydroxyvaline (NMeHV) and 3-hydroxy-6-methoxyquinoline-2-carboxylic acid (3HMQA) in the case of luzopeptine, *N*-methylnorcoronamic acid (MNCA) found in SW-163 or pipecolate (Pip) found in sandramycin.

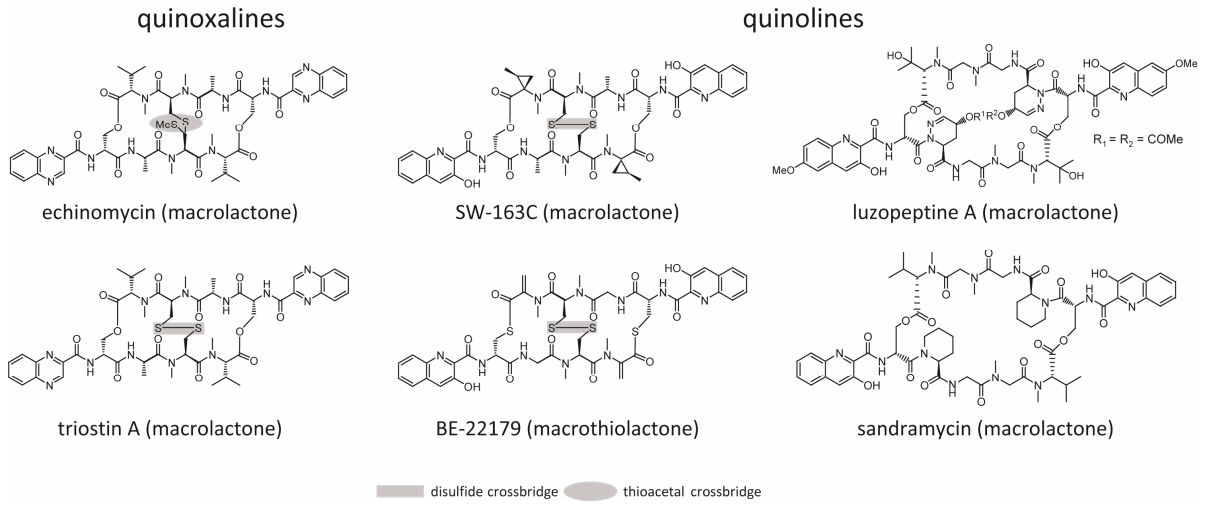



**Figure 1.13:** The class of chromodepsipeptides subdivided into the groups of quinoxalines and quinolines sharing a common peptidic scaffold and a pseudosymmetrical structure. The classification is based on the *N*-terminally attached chromophore moiety. Quinoxalines: echinomycin and triostin A. Quinolines: SW-163C, BE-22179, luzopeptine A and sandramycin. Intramolecular crossbridges, *e.g.* disulfide- and thioacetal-linkages, are accentuated in grey.

Thiocoraline itself is a twofold-symmetric bicyclic octathiodepsipeptide in which the *N*-termini of the two oligopeptide chains are capped with the chromophore moiety 3-hydroxyquinoline-2-carboxylic acid (3HQA) acting as an intercalating group (Figure 1.14 A).[68,134] Thiocoraline has been isolated from cultures of *Micromonospora* sp. L13-ACM2-092 and *Micromonospora* ML1, collected from the Mozambique strait.[68] The two symmetrical halves consisting of 3HQA-D-$Cys_1$-$Gly_2$-*N*-Me-L-$Cys_3$-*N*,*S*-dimethyl-L-$Cys_4$ are linked together through two thioester bonds between the *N*-terminal D-$Cys_1$ residue of one tetrapeptide and the *N*,*S*-dimethyl-L-$Cys_4$ of the tetrapeptide. An intramolecular disulfide crossbridge from L-$Cys_3$ residues leads to a further structural rigidification of this unique macrothiolactone. Thiocoraline shares

the D-configured *N*-terminal amino acid involved in macrocyclization with all known chromodepsipeptides, whilst the thioester bond is unique to thiocoraline and BE-22179, representing a novel class of thioesterase-mediated side-chain linkage.[20,135] The crystal structure of thiocoraline reveals the molecules to be docked in pairs with the quinoline moieties of the neighbouring units being stacked due to π-π interactions. The thiocoraline dimers form infinite columns, adopting an antiparallel helix-like orientation which resembles the mode of DNA-bisintercalation (Figure 1.14 B).[134]

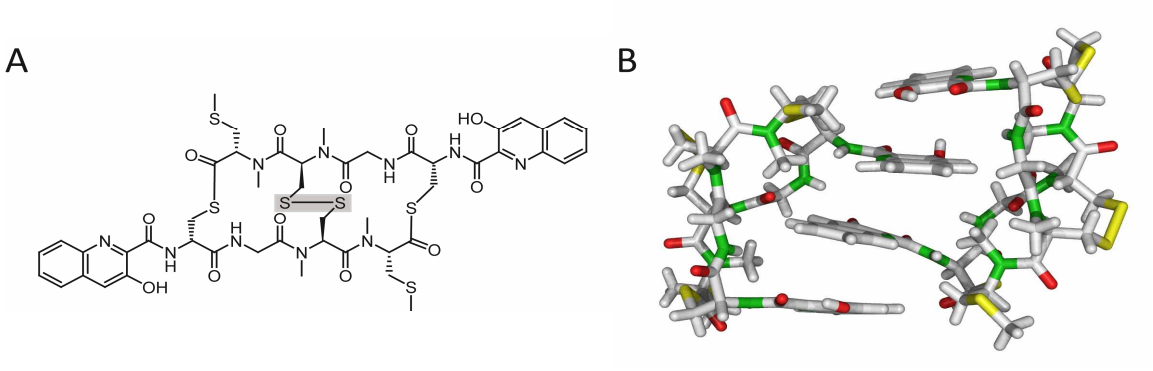



**Figure 1.14:** Structural analysis of the macrocyclic chromodepsipeptide thiocoraline. A) Chemical structure of thiocoraline. The symmetrical chromophore-capped tetrapeptides are linked together via two thioester bonds. The intramolecular disulfide bridge enhances the structural rigidity of the macrocyclic peptide. B) Crystal structure of a stacked pair of thiocoraline molecules in the asymmetric unit of the unit cell (CIF code: 629608). π-π interactions of the chromophore moieties induce an antiparallel helix-like conformation and lead to the reciprocal docking of two molecules. Oxygen, nitrogen and sulphur atoms are accentuated in red, green and yellow, respectively.

### 1.6.2 Biosynthesis of Thiocoraline

The initial identification of a gene cluster required for the biosynthesis of thiocoraline was based on insertional activation of NRPS adenylation domains in the genome of *Micromonospora* sp. ML1.[136] Two inactivations generated thiocoraline nonproducing mutants, indicating that these A-domains are involved in thiocoraline assembly. Sequencing of the thiocoraline gene cluster revealed a 64.9 kbp region, harbouring 36 CDS (Figure 1.15). Within the thiocoraline operon two dimodular NRPSs, TioR (277 kDa) and TioS (346 kDa) are encoded which mediate oligopeptide assembly. TioR contains an E-domain responsible for the epimerisation of the *N*-terminal L-Cys residue. TioS contains two *N*-methyltransferase (MT) domains catalyzing the *N*-methylation of L-$Cys_3$ and L-$Cys_4$ as well as a *C*-terminal thioesterase domain. Analysis of the specificity-conferring residues revealed the first adenylation domain to putatively activate L-Ser. $A_2$ is predicted to activate Gly and the cognate subtrate for $A_3$ is predicted to be L-Cys. The fourth adenylation domain $A_4$ is proposed to activate L-Val. The two

additional modular NRPSs, TioY and TioZ are postulated to assemble a tripeptide which might be involved in quorum sensing regulation of thiocoraline biosynthesis or interspecies communication during pathogen-host relations in aquatic ecosystems.[137-138] Mutational inactivation of TioY and TioZ, resulting in thiocoraline nonproducing mutants, confirmed both synthetases to be essential for thiocoraline production.

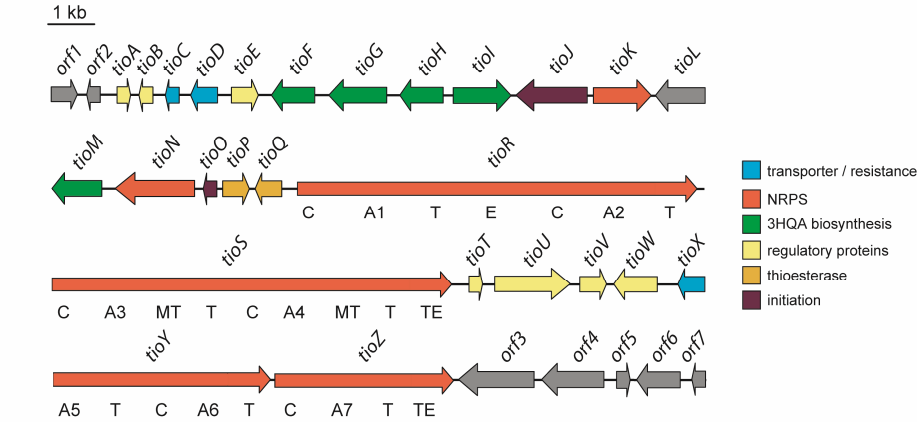



**Figure 1.15:** Schematic overview of the thiocoraline biosynthetic gene cluster. Putative functions of the proteins encoded within the operon are based on BLAST-analysis and are given in the figure. The NRPSs TioR and TioS are responsible for the assembly of thiocoraline and the genes accentuated in green are predicted to be involved in the biosynthesis of the chromophore moiety 3HQA. Additional genes encode proteins involved in thiocoraline secretion and regulation of secondary metabolite biosynthesis.

Initiation of thiocoraline biosynthesis is most likely to be carried out by the lone-standing 3HQA-AMP ligase TioJ, which activates 3HQA as an AMP-derivative which is transferred onto the ppan cofactor of the lone-standing ArCP TioO (Figure 1.16). Condensation of the ArCP bound 3HQA and L-$Cys_1$ is subsequently mediated by the *N*-terminal C-domain of the initiation module of TioR and chain elongation is commenced. Bioinformatic analysis of the gene cluster revealed a set of five genes *tioF*/*L*/*G*/*H*/*I* to encode enzymes associated with the biosynthesis of the chromophore moiety 3HQA starting from L-Trp. The first enzyme of this putative pathway to be characterized biochemically *in vitro* was the tryptophan-2,3-dioxygenase TioF which was shown to irreversibly catalyze the conversion of L-Trp to *N*-formylkynurenine.[139] Apart from core-elements required for the NRPS-based assembly of thiocoraline, two ABC-type transporters TioC/D, a set of regulatory proteins TioA/B/E/T/U/V/W and two lone-standing thioesterases TioP/Q are encoded within the cluster. Initially, it was proposed that the enzyme encoded by *tioX* is involved in the *S*-methylation of the *N*,*S*-dimethyl-L-$Cys_4$ residue as it shares 39% sequential identity to type I glyoxylases. This hypothesis was recently disproved and it was shown by

crystal structure analysis that TioX shares the bleomycin resistance protein fold and is involved in thiocoraline resistance and secretion.[140]

Due to the fact that the number of amino acids found within the product does not correlate with the total number of adenylation domains, an iterative mechanism of biosynthesis, following the described *backward* logic is proposed (Figure 1.16).[112] This mechanism requires the modular template to be used twice and the *C*-terminal TE-domain (TioS TE) to catalyze the cyclodimerization of two symmetrical tetrapeptidyl-substrates. In the first step the PCP-bound tetrapeptide is transferred onto the active-site serine of TioS TE, converting the energetically labile thioester to a more stable oxoester. The tetrapeptide remains bound to the TE as an acyl-O-TE oxoester until a second tetrapeptidyl-substrate is immobilized on the adjacent PCP. Following the *backward* mechansism, the D-$Cys_1$ sulfhydryl-group of the PCP-bound intermediate performs a nucleophilic attack onto the acyl-O-TE oxoester, resulting in a PCP-bound linear octathiodepsipeptide. This intermediate is subsequently transferred to the active-site serine of the *C*-terminal TE-domain, cyclized and released from the enzymatic template as a macrothiolactone.

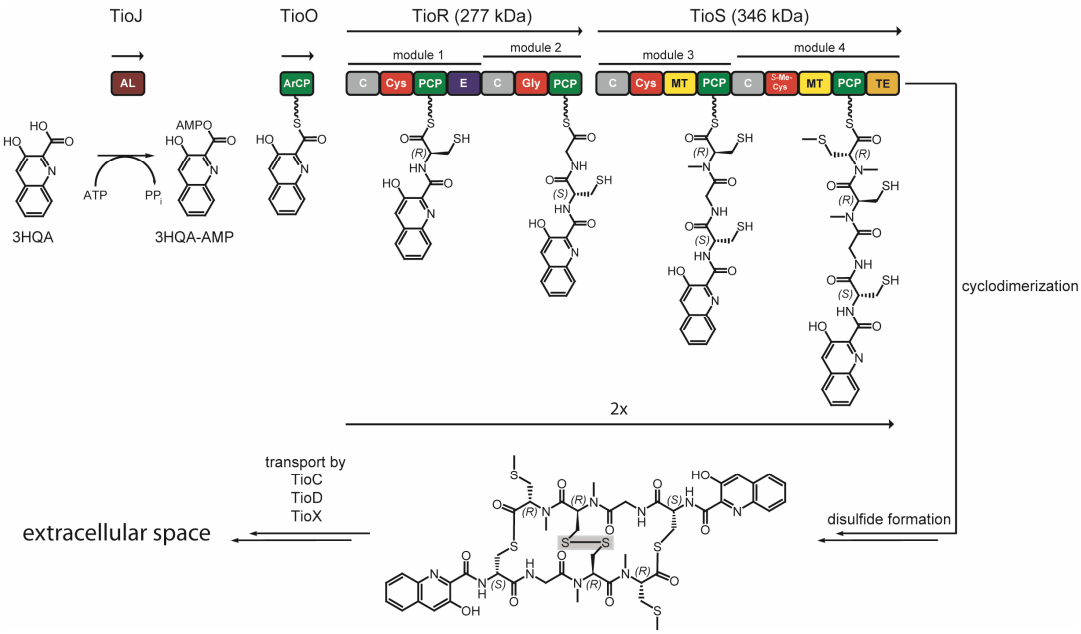



**Figure 1.16:** Nonribosomal assembly of thiocoraline by the tetramodular assembly line consisting of TioR and TioS. Initiation of thiocoraline biosynthesis is mediated by TioJ and TioO. As the number of amino acids found in the assembled product does not correlate with the four adenylation domains found in the two peptide synthetases, an iterative mechanism for thiocoraline biosynthesis is proposed. The modular template is used twice and the *C*-terminal thioesterase is mediating ligation and subsequent macrothiolactonization of two identical linear chromophore-capped tetrapeptides. In this model the transport of thiocoraline is mediated by TioC/D/X. The timing of disulfide formation remains to be elucidated.

### 1.6.3 Mode of Action of Thiocoraline

Thiocoraline is licenced to the pharmaceutical company PharmaMar (Tres Cantos, Spain) and is currently undergoing preclinical trials phase II. Thiocoraline shows potent antibacterial activity against Gram-positive bacteria and a wide spectrum of antiproliferative activity against various cancer cell lines *in vitro*.[141-142] Incubation of LoVo- and SW620 human colon cancer cells with thiocoraline induces an inhibition of cell-cycle progression and causes cell-stasis in the G1 phase of the native cell-cycle. In addition, thiocoraline decelerates the transition from the S-phase to the G2/M-phase. The pharmaceutical interest is mainly focused on the antiproliferative activity of thiocoraline towards colon cancer and metastased non-small cell lung cancer.[143] Based on the structure of thiocoraline it was postulated that intercalation into duplex DNA causes DNA topoisomerase II to induce duplex DNA breakage. DNA-unwinding assays in the presence of DNA topoisomerase II disproved this mode of action (MOA) as neither inhibition of the enzyme nor DNA breakage could be observed. In contrast, primer-extension assays confirmed the inhibition of DNA replication by DNA-polymerase $\alpha$ *in vitro*. It was postulated that thiocoraline intercalates into specific sequences of duplex DNA and induces the dissociation of DNA-polymerase $\alpha$ from the primer/template complex leading to the termination of DNA replication.[141] The DNA-intercalative properties of the macrocyclic product have been confirmed via DNA-melting assays and it was shown that duplex DNA is stabilized upon incubation with thiocoraline.[134] Recently, a structural basis for the mode of action of thiocoraline has been established through molecular dynamics simulation of thiocoraline bisintercalating into duplex DNA.[134] Thiocoraline is shown to adopt a U-shaped conformation and to bind to the minor groove of GC-rich sequences, especially those encompassing a central CpG step presumably leading to an inhibition of DNA polymerase $\alpha$. The planar chromophore moiety 3HQA ensures a tricyclic hydrogen-bonded conformation and facilitates DNA-bisintercalation. In addition, hydrogen-bonds are formed between the NH- and carbonyl groups of glycine and the N3 and exocyclic amino group of guanine.[134]

A main drawback of thiocoraline is its half-life and the immediate metabolization of the compound in human serum.[144] Organic synthesis of the aza- and oxathiocoraline-class has led to compounds with increased physico-chemical stability and an increased half-life in human plasma from currently 4 h to clinically applicable time spans.[145] The development of *in vitro*-approaches to obtain analogues of thiocoraline with improved physico-chemical stability and to circumvent low yield organic synthesis is crucial for the generation of potential therapeutic applications based on this class of compounds.

### 1.6.4 Bacterial Siderophores

Bacterial growth is strongly influenced by the availability of iron as an essential trace element employed as cofactor.[146] The fact that the bioavailability of iron is challenging for most microorganisms, since it is mostly found in the insoluble Fe(III) (ferric iron) redox state, forming insoluble $Fe(OH)_3$ complexes, has led to the evolutionary development of highly efficient iron uptake systems. In response to iron starvation many microorganisms produce and secrete iron-scavenging compounds (generally < 1 kDa) termed siderophores, with a high affinity for ferric iron ($K_f$ = $10^{22}$ -$10^{49}$ $M^{-1}$).[147] Most siderophores coordinate iron via six donor atoms as octahedral complexes in an ferric iron:siderophore-ratio of 1:1. But also ligand:iron-ratios of 3:2, as in the case of rhodoturolic acid or 2:1, in the case of pyochelin, have been reported if the coordination sphere of the ligand:iron-complex exhibits unoccupied coordination sites.[148-149] These sites can be either occupied by alternative oxygen donors such as water molecules or by additional ligands. The siderophore-based iron acquisition is distributed widely among different genera of prokaryotic and eukaryotic microbes and is also employed by fungi (*e.g. Ustilago maydis*).[150] After extracellular binding of iron the siderophores are reimported into the cell after recognition by specific receptors and iron is released from the chelator-complex to be subsequently channelled to the intracellular targets.[151-153] Siderophores are structurally diverse natural products and several hundred members have been isolated from microbial and fungal sources. Siderophores can be subdivided into three main classes, depending on the chemical nature of iron-chelating moieties required for octahedral coordination of iron. These can either be catecholates (phenolates), hydroxamates, (α-hydroxy-)carboxylates or combinations hereof in mixed-type siderophores. The most commonly known catecholate siderophores are the triscatecholate siderophores enterobactin, bacillibactin and the glucosylated salmochelin, all containing three 2,3-DHB-moities for ferric iron chelation (Figure 1.17).[154-156] Triscatecholate siderophores exhibit the highest stability constants for the siderophor:metal complex with $Fe(enterobactin)^{3-}$ at $10^{49}$ $M^{-1}$ and $Fe(bacillibactin)^{3-}$ at $10^{47}$ $M^{-1}$.[157] Trishydroxamate siderophores contain hydroxamic acid moieties for iron chelation usually consisting of hydroxylated and/or formylated and acetylated lysine or ornithine residues. An increasing number of marine amphiphilic siderophores exemplified by the group of amphibactins and aquachelins has been isolated, containing these functional groups (Figure 1.17).[158-159] Carboxylate-type siderophores generally utilize α-hydroxycarboxylic acid as bidentate chelating group for the coordination of ferric iron. In the case of staphyloferrin A two citrate groups provide the α-hydroxycarboxylates for iron coordination (Figure 1.17).[160] A representative overview of the different siderophore classes based on the functional groups involved in the chelation of iron is given in Figure 1.17.

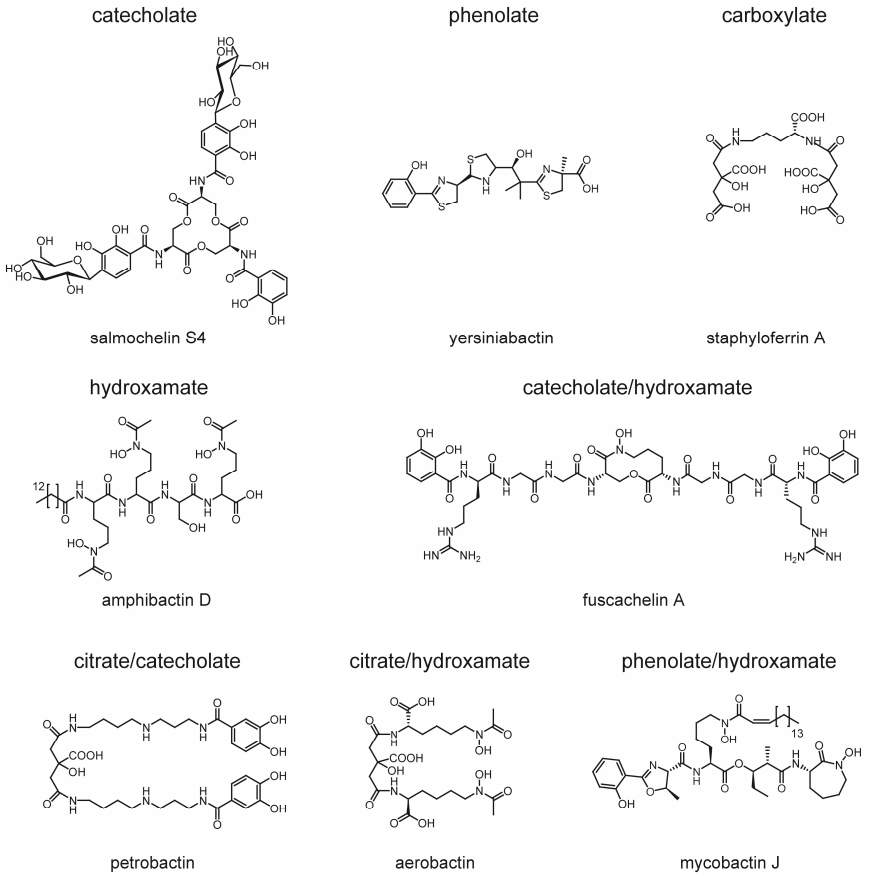



**Figure 1.17:** Structures of representative siderophores. The classification is based on the functional groups involved in octahedral coordination of ferric iron. Siderophores are: salmochelin (*Salmonella enterica*), yersiniabactin (*Yersinia pestis*), staphyloferrin A (*Staphylococcus* spp.), amphibactin D (*Vibrio* sp. R-10), fuscachelin A (*Thermobifida fusca* YX), petrobactin (*Bacillus anthracis*), aerobactin (*Shigella flexneri* 5 str. 8401) and mycobactin J (*Mycobacterium tuberculosis*).[37,161-163]

### 1.6.5 Biosynthesis of NRPS-Depedent and NRPS-Independent Siderophores

In addition to the functional groups involved in iron chelation, siderophores can also be classified based on their biosynthetic origin. The first and most extensively studied class of siderophores is produced by NRPSs following linear, nonlinear and iterative assembly logics. The second class is known as NRPS-independent siderophores (NIS) and involves a novel family of synthetases, represented by IucA and IucC, responsible for aerobactin (*Escherichia coli* K-12, *Shigella flexneri* 5 str. 8401) biosynthesis.[99,165] Siderophores of NRPS-independent origin encompass desferrioxamine E (*Streptomyces coelicolor* M145), putrebactin (*Shewanella putrefaciens*) and further compounds.[166-167] In both cases the biosynthetic genes

of these secondary metabolites are usually clustered within one operon showing coordinated transcriptional regulation.[168]

Nonribosomal assembly of siderophores and modification of amino acids by tailoring enzymes is exemplified by the synthesis of the trishydroxamate-type siderophore coelichelin, isolated from cultures of *S. coelicolor* A3(2) (Figure 1.18). Initiation of coelichelin biosynthesis is mediated by the FAD-dependent monooxygenase CchB, which catalyzes the hydroxylation of L-ornithine.[98] This hydroxylation of the δ-amino group is common to a variety of NRPS-derived natural products and is usually catalyzed by external FAD-dependent monooxygenases. The most recently characterized FAD-dependent monooxygenases involved in the generation of building blocks for NRPS-derived siderophores are PvdA (*Pseudomonas aeruginosa*) and CchB (*S. coelicolor* A3(2)), encoded in the gene clusters responsible for pyoverdine or coelichelin biosynthesis.[98,169] δ-*N*-hydroxy-L-ornithine is either directly incorporated into the oligopeptide by the third module of the synthetase CchH or formylated at the δ-amino group by the putative formyltransferase CchA. δ-amino group formylation is also observed during the synthesis of pyoverdine and it is assumed that enzymes sharing a high degree of sequential identity with methionyl-tRNA-formyltransferases catalyze this modification. The cognate formyl-group donor is postulated to be $N^{10}$-formyltetrahydrofolate ($N^{10}$-f$H_4$F). Coelichelin assembly, catalyzed by the trimodular NRPS CchH follows a nonlinear logic due to the presence of four amino acids within the readily synthesized siderophore. It is postulated that the first step of coelichelin biosynthesis is the activation and covalent tethering of δ-*N*-formyl-δ-*N*-hydroxy-L-ornithine (hfOrn) by the first module of CchH. In the next step $hfOrn_1$ is condensed with $Thr_2$. Subsequently, the dipeptide is connected to $hOrn_3$ via isopeptide linkage. In the ultimate step the α-amino group of $hOrn_3$ attacks the PCP-bound $hfOrn_4$ in a nucleophilic manner. It is postulated that the PCP-bound tetrapeptide is hydrolytically released from the enzymatic template by the α/β-hydrolase-type enzyme CchJ, as no TE-domain could be identified within the synthetase CchH.

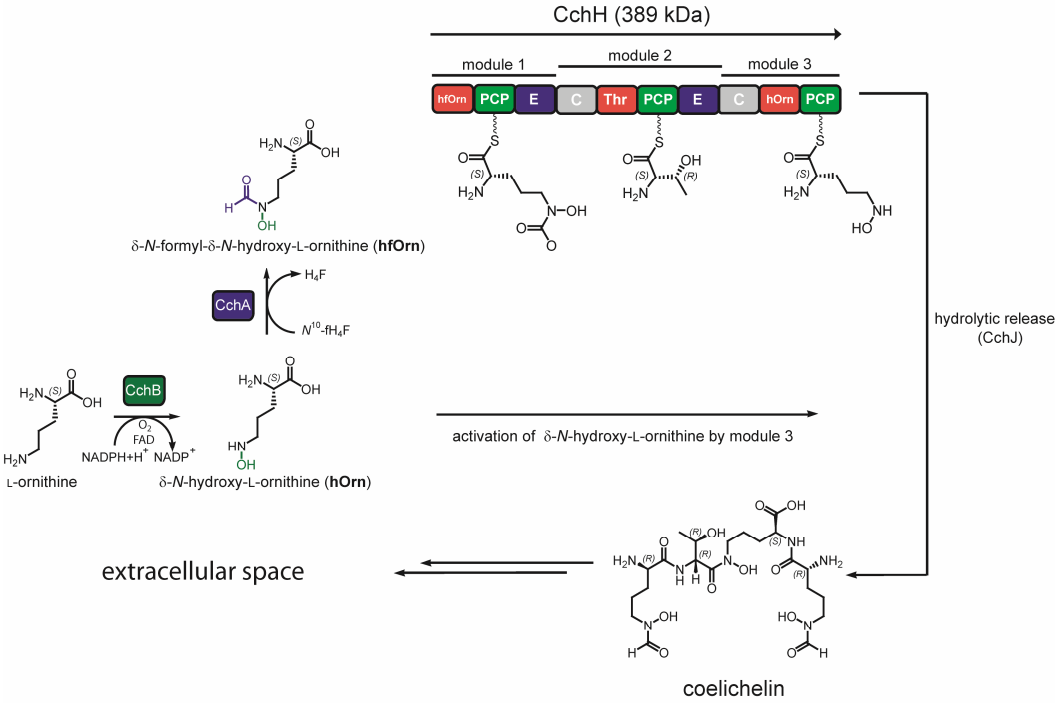



**Figure 1.18:** The postulated model for the nonlinear nonribosomal assembly of the tetrapeptide siderophore coelichelin. The FAD-dependent monooxygenase CchB provides hOrn which is either formylated by the putative formyltransferase CchA or directly activated by module 3 of the synthetase CchH. Hydrolytic release of the tetrapeptide is catalyzed by CchJ.

The second class of siderophores is synthesized following a NRPS-independent pathway. The first NIS to be discovered was the mixed citrate/hydroxamate-type siderophore aerobactin, which is produced in response to iron-starvation by several Gram-negative bacteria including *E. coli*, *Shigella*, *Yersinia* and *Salmonella* species.[170-171] The gene cluster governing aerobactin assembly is comprised of four genes and was initially identified on the pColV-K30 plasmid of *E. coli* (Figure 1.19).[171-172] Analysis of the cluster revealed *iucD* to encode a FAD-dependent monooxygenase that catalyzes ε-amino group hydroxylation of L-lysine residues in analogy to CchB, catalyzing the δ-amino group hydroxylation of L-ornithine.[173] Aerobactin biosynthesis requires an additional tailoring step for the generation of ε-*N*-acetyl-ε-*N*-L-hydroxylysine (haLys). This acetylation of ε-*N*-L-hydroxylysine (hLys) is mediated by the acetyltransferase IucB, which catalyzes acetyl-transfer from acetyl-CoA to the hydroxylamino group.[174] It was shown that IucB transfers the acetyl-moiety of acetyl-CoA to a variety of hydroxylamines with hLys being the preferred substrate. In a biosynthetic model for aerobactin production based on genetic studies, it is postulated that the synthetase IucA mediates the acylation of the α-amino group of haLys with a prochiral carboxyl group of citric acid. Subsequently, the synthetase IucC catalyzes a second acylation of citric acid with haLys to give rise to aerobactin (Figure 1.19). It is assumed that the IucA- and

IucC-mediated acylations of citric acid require nucleotide triphosphates (NTPs) as cosubstrates for the activation of carboxyl groups. Recently, isolated and biochemically characterized NIS synthetases were shown to require ATP for the generation of desferrioxamines or putrebactin and it is expected that IucA/IucD utilize ATP as cosubstrates as well.[165,175] IucA and IucC are considered the founding members of NIS synthetases as every NIS biosynthetic pathway involves at least one synthetase with a high degree of sequential identity to IucA/IucC.[176]

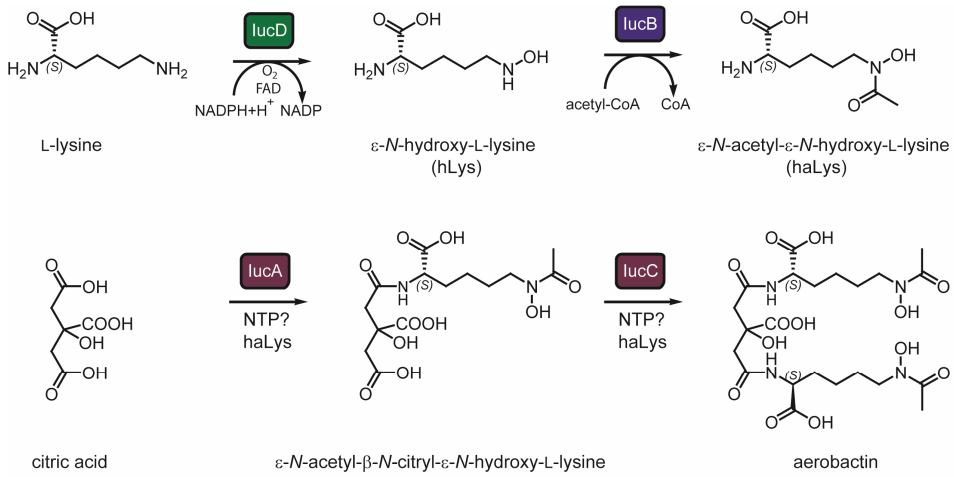



**Figure 1.19:** The postulated pathway for aerobactin biosynthesis. The FAD-dependent monooxygenase IucD mediates ε-amino group hydroxylation of L-Lys. Acetylation of hLys is carried out by the acetyltransferase IucB and gives rise to haLys, which is subsequently condensed with a carboxyl group of citric acid by the NIS synthetase IucA. IucC catalyzes a second condensation reaction that ultimately affords the siderophore aerobactin.

## 1.7 Rational Strategies for the Generation of Structural Diversity

The prevalence of bacterial resistance to contemporary antibiotics represents a challenge, which has to be addressed by the constant development of novel antibiotics. Three main strategies have been developed over the last centuries for the generation of structural and functional diversity of natural products: total organochemical synthesis of pharmacologically active compounds, semi-synthetic approaches for the generation of derivatives based on bioactive lead structures or the identification and isolation of novel natural products. Within this chapter two main strategies are addressed:

I) Chemoenzymatic synthesis of NRP analogues based on the original lead structure

II) Natural product discovery via genome mining

### 1.7.1 Chemoenzymatic Synthesis of NRP Derivatives

Insights into the nonribosomal biosynthesis of peptides enabled a chemoenzymatic approach for the generation of NRP analogues by combining organic synthesis and enzymatic mechanisms. The catalytic unit utilized in this approach is the TE-domain usually located in the *C*-terminal module of the ultimate NRPS. TE-domains remain catalytically active units when separated from the enzymatic template and can be incubated with synthetic substrates for the evaluation of their inherent biocombinatorial potential.[118] These substrates are synthesized via solid-phase peptide synthesis (SPPS), substituting the entire enzymatic machinery.[119,177] Recognition and cyclization of the synthetic substrate requires the *C*-terminal activation of the oligopeptide as a thioester, imitating the physiological situation of the naturally occurring PCP-ppan-bound substrate.[178] *N*-acetylcysteamine (SNAC) activation proved to be advantageous for this biomimetic approach but thiophenol- or CoA-activations have also been successfully applied.[118,179] The incubation of the SrfA-C TE or the Tyc TE with SNAC-activated peptidylthioesters afforded chemoenzymatically generated surfactin- or tyrocidine-analogues *in vitro*.[53,177] A limiting factor of the biomimetic approach results from the leaving group tolerance of several TE-domains. The *apo*-PCP-TE-didomain of the fengycin system had to be posttranslationally converted into the corresponding *holo*-PCP-TE with peptidyl-CoA substrates. Incubation of the recombinant Fen PCP-TE with peptidyl-CoA substrates and the PPtase Sfp (*Bacillus subtilis*) gave rise to macrocyclic products.[179] As the PCP-domain remains in the *holo*-form after *in vitro* phosphopantetheinylation, this method is constricted to a single turnover mechanism. More recent studies focused on the generation of daptomycin and A54145 derivatives and hybrid molecules through enzymatic cyclization of linear thioesters by the TE-domains from A54145 and daptomycin NRPSs (Figure 1.20).[180] Structure-activity-relationship (SAR) studies were conducted with a set of daptomycin

analogues including an alternative macrolactam to evaluate the importance of several residues of the ten-membered macrocyclic ring. Summarizing this chemoenzymatic approach, it can be stated that this method offers the possibility to rapidly construct NRP analogues that can be investigated towards SAR. As the main drawback of this method *in vitro* is the low quantity of derivatives generated, scientists of the post-genomic era set out to exploit the known biosynthetic machinery *in vivo* and to discover new natural product biosynthetic chemistry by genome mining.

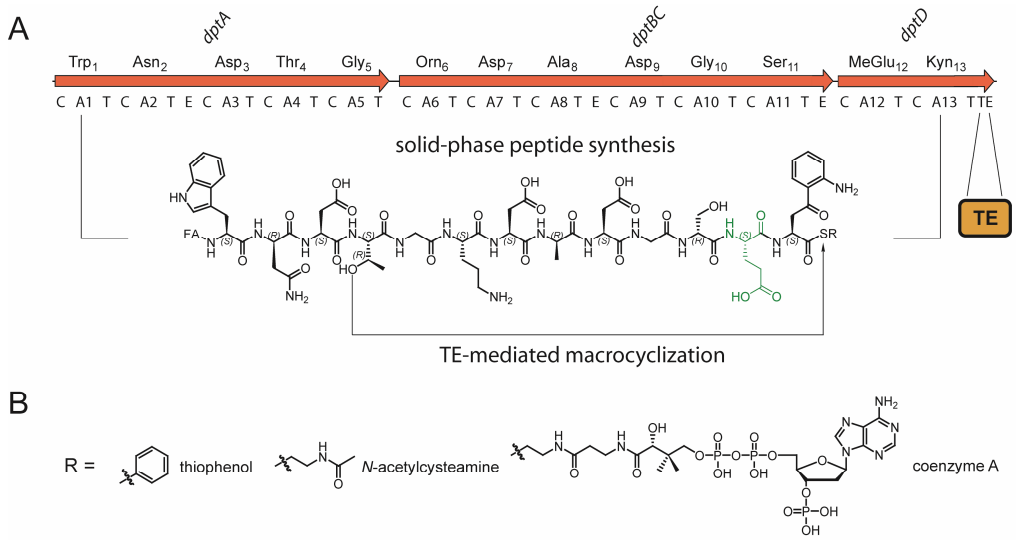



**Figure 1.20:** A) *In vitro*-chemoenzymatic synthesis of daptomycin derivatives employing linear, *C*-terminally activated peptidyl thioesters substituting the entire assembly line. The catalytic unit responsible for macrolactonization is the DptD TE-domain, dissected from the enzymatic template. The modified residue is accentuated in green. B) Chemical structures of the *C*-terminal leaving groups.

### 1.7.2 Natural Product Discovery via Genome Mining

Natural product discovery is traditionally based on bioassay-guided fractionation of extracts from natural sources and subsequent isolation of the bioactive compound leading to the identification of a multitude of antibacterial, antifungal, anticancer and immunosuppressive agents.[2] One of the first rational strategies for the discovery of natural products was the systematic cultivation of one organism under several conditions giving rise to the one strain/many compounds (OSMAC) approach.[181] It was shown that the environment and the cultivation conditions strongly influence the secondary metabolome of the target strain, revealing the potential of the strain to produce natural products. Application of this approach to *Streptomyces* sp. Gö40/10 led to the discovery of various compounds upon systematic variation of the cultivation parameters.[181] With the advance in sequencing technologies, coming from

whole genome shotgun sequencing to high-throughput pyrosequencing proliferating over 1000 sequenced and annotated microbial genomes, a multitude of gene clusters related to natural product biosynthesis was revealed (source: Genomes Online Database - GOLD).[182-183] This quantity of information in combination with a substantial increase in the understanding of natural product biosynthetic mechanisms has paved the way for the mining of genomes for bioactive compounds.[25] Genome mining describes a methodology for the isolation of natural products predicted solely from the genome sequence of the target organism. As many microbial natural products, *e.g.* polyketides (PKs) and NRPs, are assembled by multimodular synthases and synthetases, bioinformatic sequence analysis tools have been developed for the prediction of module organization and substrate specificity.[25,184-185] The first microbial organism to be extensively analyzed towards the production of unknown secondary metabolites was *Streptomyces coelicolor* A3(2)[186]. Within the genome of *S. coelicolor* A3(2), several gene clusters were identified encoding biosynthetic pathways for natural product assembly.[187] Many of these clusters are considered *cryptic* as they are not associated with known metabolites. Mining of the *S. coelicolor* genome led to the discovery of several novel compounds now associated with formerly cryptic gene clusters.[38,188-190] Throughout the years, several concepts for the mining of microbial genomes have been developed combining the bioinformatic analysis of the sequenced genome, identification of cryptic gene clusters and identification of conditions leading to gene cluster expression either in the original strain or a heterologous host. One approach solely relies on the prediction of the physico-chemical properties of the target compound, including bioactivity, and has been successfully applied for the identification of salinilactam A (*Salinispora tropica*).[191] Sequence analysis of *Salinispora tropica* revealed a modular PKS gene cluster putatively encoding a lysine-primed polyene macrolactam. Subsequently, salinilactam A was identified based on the characteristic polyene UV-absorption properties and structurally elucidated via NMR and MS analysis (Figure 1.21 A). An alternative approach is represented by the *in vitro* reconstitution of natural product biosynthesis. This tedious approach circumvents the problem of gene clusters not expressed *in vivo*, as it uncouples the biosynthetic enzymes from native regulatory mechanisms, and requires the recombinant production of all biosynthetic enzymes and the corresponding cofactors and predicted substrates. Application of this method led to the discovery of the two-component lantibiotic haloduracin (*Bacillus halodurans* C-125) by incubation of the biosynthetic enzymes with the ribosomally synthesized substrates *in vitro*.[192]

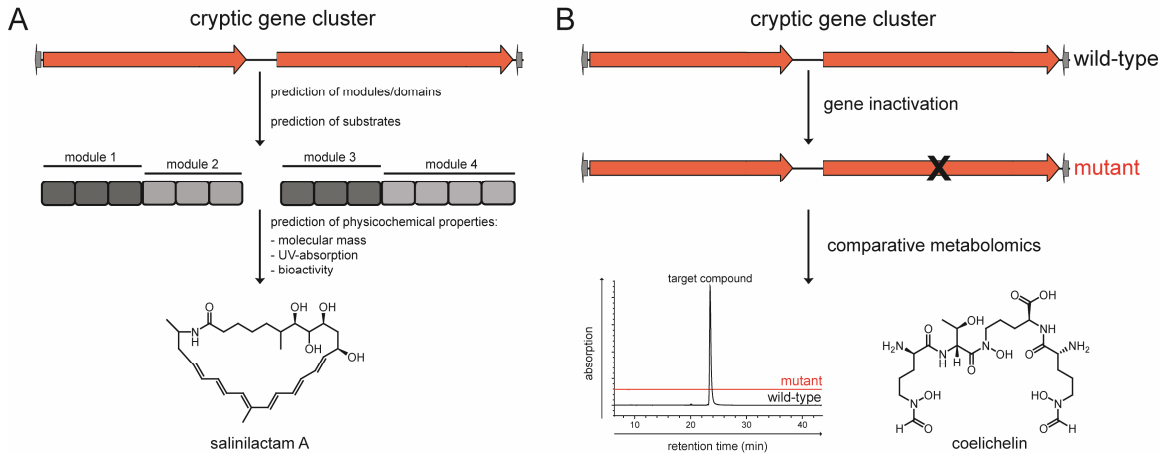



**Figure 1.21:** Genome mining strategies applied for the discovery of salinilactam A and coelichelin. A) Outline of the genome mining concept relying on the prediction of the physico-chemical properties (*e.g.* molecular mass, UV-absorption or bioactivity) of the cryptic natural product. B) The gene knockout/comparative metabolic profiling concept. Genes located in the cryptic gene cluster are inactivated and the cryptic natural product is identified via HPLC-comparison of the wild-type and mutant metabolome.

The comparative metabolomic profiling approach has been applied for the identification of genes responsible for the biosynthesis of natural products, but it can also be carried out for natural product discovery (Figure 1.21 B). The inactivation of genes located in cryptic gene clusters and subsequent comparative metabolome studies of the wild-type and the mutant strain by LCMS can result in the identification of cryptic metabolites and has been successfully applied for the discovery of the hydroxamate-type siderophore coelichelin.[38] In summary, this method represents an effective tool for natural product discovery as it does not require secondary metabolite structure prediction but is limited by the genetic accessibility of the target strain. An additional limitation is the quantity of cryptic natural product produced, as the change in the metabolite profile of the mutant and the wild-type strain has to be detectable.

A genome mining concept similar to the comparative metabolomic profiling approach involves the heterologous expression of cryptic biosynthetic gene clusters in combination with metabolome profiling. This method requires the insertion of the cryptic gene cluster into the genome of a nonproducing host organism via recombination techniques or the transformation of the host with the plasmid- or cosmid-encoded cryptic gene cluster. Comparison of the metabolite profiles of the host containing the gene cluster with the wild-type host affords the identification of the cryptic natural product as it is absent in cultures of the wild-type strain. Application of this concept has led to the discovery of methylenomycin furans (*S. coelicolor*).[189] The expression of a biosynthetic gene cluster in a heterologous host offers the

opportunity to perform pathway engineering *in vivo* for the generation of natural product derivatives or for the functional investigation of the biosynthetic enzymes. Transformation of *E. coli* with the cosmid-encoded andrimid or echinomycin biosynthetic gene clusters resulted in the heterologous production of the antibiotic andrimid or the antibiotic compound echinomycin and granted insights into the biosynthetic mechanisms leading to these compounds.[193-195]

A more general strategy for natural product discovery is designated genomisotopic approach, which has been successfully applied for the identification and isolation of the NRPS-derived cyclolipopeptide orfamide A (*Pseudomonas fluorescens* Pf-5).[196] The genomisotopic method combines bioinformatic prediction of substrate specificity for the biosynthetic pathway with the incorporation of a stable-isotope labeled precursor into the assembled product (Figure 1.22). In the course of orfamide A isolation, the NRPS-catalyzed incorporation of four L-Leu residues into the lipopeptide was postulated based on A-domain specificity prediction. In order to identify the cryptic product, $^{15}$N-labeled L-Leu was fed to cultures of *Pseudomonas fluorescens* Pf-5 and culture extracts and HPLC fractions were subjected to $^{1}$H-$^{15}$N-HMBC NMR experiments for the identification of metabolites containing the isotope label.

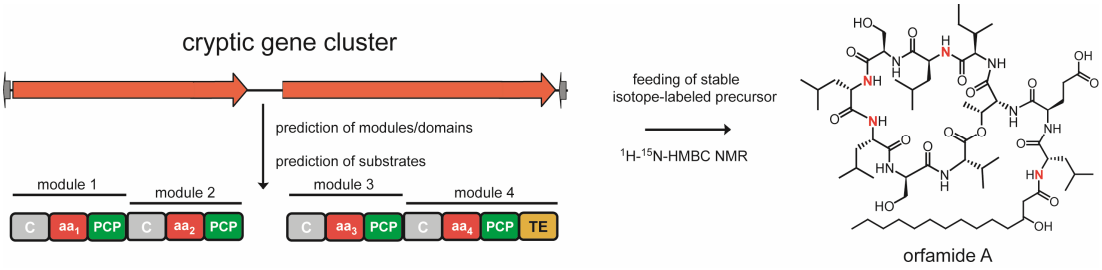



**Figure 1.22:** The genomisotopic approach for the identification of the cryptic NRPS-derived cyclic lipopeptide orfamide A. Bioinformatic analysis of A-domain specificities predicted the incorporation of four L-Leu residues into the octapeptide. $^{15}$N-L-Leu was fed to the cultures of *Pseudomonas fluorescens* Pf-5 and subsequent $^{1}$H-$^{15}$N-HMBC NMR analysis led to the identification and isolation of the target compound. Potential incorporation sites of the isotope-label are accentuated in red.

The accurate prediction of substrate specificity was found to be crucial for successful mining and structural prediction and is the basis of the genomisotopic methodology.[196] In addition, the time and amount of precursor feeding has to be optimized for the sufficient incorporation of the label for NMR experiments. In the course of the study presented herein, the genomisotopic approach was improved by combining transcriptome analysis of the target strain, bioinformatic A-domain specificity prediction and an unprecedented radio-LCMS setup for the establishment of a highly sensitive genome mining method for NRPS-derived natural product discovery.

## 2. Objectives of this Study

In order to obtain insights into the biosynthetic origin of chromodepsipeptides, the biochemical characterization of the thioesterase domain derived from the thiocoraline NRPS TioS was one objective of this study. To address this objective, suitable tetrapeptidyl thioesters, resembling the repetitive building blocks of thiocoraline, should be synthesized via solid-phase peptide synthesis (SPPS) to investigate the unprecedented macrothiolactonization reaction and the substrate specificity. The biocombinatorial potential of TioS PCP-TE should be evaluated for the generation of chromodepsipeptide analogs with altered chemical structures. These chemoenzymatically generated macrocycles should be isolated and investigated towards DNA-bisintercalation activity. Within this context, a DNA-bisintercalation assay should be established to compare DNA-stabilization induced by the analogs with DNA-stabilization induced by authentic thiocoraline.

A second objective of this study was the development of a general genome mining strategy for the discovery of cryptic NRPS-derived natural products. This radio-LCMS-guided genome mining approach should be applied to the sequenced and annotated strain *Saccharopolyspora erythraea* NRRL 23338. Prior to application of the genome mining approach, bioinformatic analysis of the annotated genome should be carried out to identify potential NRPS gene clusters to be targeted with the developed methodology. The newly identified natural product should then be isolated in sufficient amounts for structure elucidation via NMR- and $MS^n$-based analytical methods and for the determination of additional physicochemical properties.

The biochemical characterization of erythrochelin biosynthesis *in vitro* was the third objective of this thesis. Erythrochelin represents a hydroxamate-type siderophore, discovered via radio-LCMS-guided genome mining of cultures of *S. erythraea*, containing two unusual δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine (L-haOrn) residues. Bioinformatic analysis of the *S. erythraea* genome should be applied for the identification of genes encoding the enzymes involved in L-haOrn biosynthesis. The corresponding enzymes should be heterologously produced and biochemically characterized *in vitro*. Evaluation of substrate specificities and kinetic parameters should result in a singular biosynthetic route for the generation of L-haOrn and enable the establishment of a biosynthetic model for erythrochelin assembly.

# 3. Materials

## 3.1 Chemicals, Enzymes, General Materials and Consumables

Chemicals and products not listed below were purchased from Sigma-Aldrich (Steinheim, Germany) in p.a. grade and were used without further purification unless noted otherwise.

**Table 3.1:** Chemicals, enzymes, general materials and consumables.

| manufacturer (*site*) | product |
|---|---|
| Agilent Technologies (*Böblingen, Germany*) | MALDI-TOF DHB-matrix |
| AnaSpec (*Fremont, USA*) | Fmoc-L-Cys-OH |
| AppliChem (*Darmstadt, Germany*) | ampicillin, kanamycin, media components |
| ARS (*Peoria, USA*) | *Saccharopolyspora erythraea* NRRL 23338 |
| Bachem (*Weil am Rhein, Germany*) | Hip-His-Leu-substrate, $N_\alpha$-Fmoc-protected amino acids, $N_\alpha$-Boc-protected amino acids |
| Biomol (*Ilvesheim, Germany*) | DTT |
| Brand (*Wertheim, Germany*) | Plastbrand PS cuvettes |
| CECT (*Valencia, Spain*) | *Micromonospora sp.* L13-ACM2-092 |
| Eppendorf (*Hamburg, Germany*) | 1.5 and 2.0 mL reaction tubes |
| Eurofins MWG Operon (*Ebersberg, Germany*) | oligonucleotides |
| Eurogentech (*Seraing, Belgium*) | agarose, electroporation cuvettes |
| EZBiolabs (*Westfield, USA*) | pBluescriptIISK(+)(*tioS PCP-TE*) |
| Fermentas (*St. Leon-Rot, Germany*) | protein size markers |
| Finnzymes (*Espoo, Finnland*) | Phusion DNA-polymerase |
| GATC (*Konstanz, Germany*) | DNA sequencing |
| GE Healthcare (*Freiburg, Germany*) | Coomassie Brilliant blue G and R250, Hi-Trap desalting columns, illustra NAP-10 columns, IPTG, size exclusion chromatography columns, yeast extract |
| Hartmann Analytic (*Braunschweig, Germany*) | $^{14}$C-L-ornithine |
| Invitrogen (*Karlsruhe, Germany*) | *E. coli* strains (BL21, TOP10) |
| Iris Biotech (*Marktredwitz, Germany*) | $N_\alpha$-Fmoc-protected amino acids |
| Macherey & Nagel (*Düren, Germany*) | $C_8$/$C_{18}$ HPLC columns (Nucleosil, Nucleodur) |
| Merck4Biosciences (*Nottingham, UK*) | 2-chlorotrityl resin, HBTU, HOBt, pet28a(+)-vector, PyBOP |
| Millipore (*Schwalbach, Germany*) | dialysis membranes (pore size: 0.025 µM), Amicon Ultra-15 concentrators |
| New England Biolabs (*Ipswitch, USA*) | Desoxyribonucleotides (dATP, dTTP, dGTP, dCTP), DNA ladders, protein size markers, restriction endonucleases |
| Novabiochem (*Bad Soden, Germany*) | $N_\alpha$-Fmoc-protected amino acids, |
| Oxoid (*Cambridge, UK*) | Agar Nr. 1, tryptone |
| Phenomenex (*Torrance, USA*) | Synergi Fusion RP-80 HPLC column |
| QIAgen (*Hilden, Germany*) | *E. coli* M15/pREP4, QIAprep Miniprep kit, QIAquick gel extraction kit, QIAquick-spin PCR purification kit, Ni-NTA-IMAC-resin, Genomic tip-20/G columns |
| Roth (*Karlsruhe, Germany*) | acrylamide solution, β-mercaptoethanol, ethidium bromide, piperidine |
| Sarstedt (*Nümbrecht, Germany*) | pipette tips, Falcon tubes (15 and 20 mL) |
| Serva (*Heidelberg, Germany*) | Bromophenol blue, Triton X-100 |
| Stratagene (*La Jolla, USA*) | *E. coli* XL1-Blue, PfuTurbo DNA polymerase |
| Sigma-Aldrich (*Steinheim, Germany*) | ACE, acetyl-CoA, FDAA, glycerol,malonyl-CoA, *N*-acetylcysteamin, oligonucleotides, SDS, TCEP, TEMED, XAD16 resin |
| Thermo Scientific (*Waltham, USA*) | Hypercarb HPLC column |

### 3.2 Equipment

**Table 3.2**: The equipment used in the course of this project.

| device | manufacturer and type |
|---|---|
| autoclave | Tuttnauer 5075 ELV, Fedegari Autoclavi SPA FVA3/A1 |
| CD-spectrometer | Jasco J-810 |
| centrifuges | Eppendorf 5415D, Eppendorf 5702R, Eppendorf 5810R, Heraeus Biofuge pico, Sorvall RC 5B Plus (SS-34, SLC-300, SLC-4000 rotors), VWR Galaxy microcentrifuge 16 |
| clean bench | Antair BSK |
| documentation system for DNA-agarose gels | Cybertech CS1 camera, Mitsubishi video copy thermo printer |
| electrophoresis | agarose gel chambers manufactured in-house (PUM, Marburg), Bio-Rad Mini-PROTEAN 3 gel chamber |
| electroporation-pulser | Bio-Rad gene pulser and pulse-controller |
| Fast protein liquid chromatography (FPLC) | Amersham Pharmacia FPLC system 250: gradient programmer GP-250, pump P-500, Uvicord optical device UV-1 ($\lambda$ = 280 nm), Uvicord control element UV-1injection valve V-7, 2-channel printer REC-101, 3-way valve PSV-100, fraction collector FRAC-100<br>Amersham Pharmacia Biotech Äktaprime |
| fluidizer | Avestin EmulsiFlex-C5 |
| french-press | SLM Aminco french-pressure cell version 5.1<br>20k Rapid-fill cell (35 mL) |
| HPLC fraction collector | Agilent AnalytFC |
| HRMS system | Thermo Fischer Scientific LTQ-FT / Agilent 1100 HPLC system |
| incubators | Köttermann 2736 |
| LC-MS system | Agilent series 1100 HPLC-system with DAD- and ESI-Quad-MS-detection, vacuum degasser, quarternary pump, auto sampler , HP-ChemStation software, optional Agilent 1100 FLD |
| lyophilization | Christ Alpha 2-4 LSC |
| MALDI-TOF | Bruker BiFlex III |
| NMR | Bruker AV600 |
| peptide synthesizer | Advanced ChemTech Apex 396 synthesizer |
| pH meter | Schott CG 840 |
| photometers | PEQLab Nanodrop ND-1000; Pharmacia Ultrospec 3000 |
| Q-TOF MS | Applied Biosystems API QStar Pulsar i |
| radioactivity flow-through detector | Berthold FlowStar LB513 / YG-40-U5M microbore cell |
| pipettes | Eppendorf Research series, Abimed Labmate |
| spectropolarimeter | Jasco J-810 |
| spectrophotometer | Beckmann Coulter DU-800 / peltier temperature controller |
| speed-Vac | Uniequip Univapo 150H |
| thermal cycler | Eppendorf Mastercycler Personal |
| thermomixer | Eppendorf Thermomixer comfort |
| vortexer | Scientific Industries VortexGenie2 |
| water deionizer | Seral Seralpur Pro90CN |

## 3.3 Plasmid Vectors

### 3.3.1 pQE60

The pQE60 vector (QIAgen), belonging to the QIAExpressionist system, was used for the cloning and expression of *tioS PCP-TE*. The vector sequence appends a *C*-terminal $His_6$-tag upon translation, allowing the isolation of the corresponding protein via Ni-NTA affinity chromatography. The vector carries two *lac*-operator recognition sequences in the promoter region enabling IPTG-induced expression of the target gene. The promoter and terminator sequence is specific for the T5 RNA polymerase. The *bla*-gene confers resistance to ampicillin up to a final concentration of 100 µg/mL.

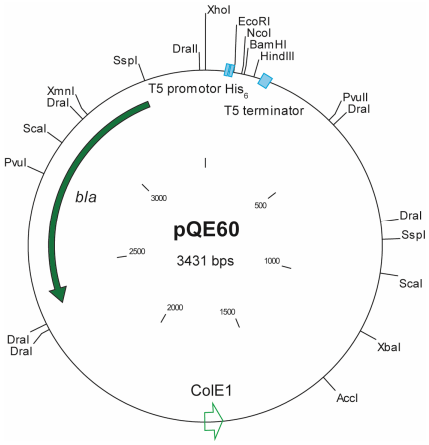



**Figure 3.1:** Physical map of pQE60.

### 3.3.2 pET28a(+)

The pET28a(+) expression vector (Novagen) was employed as a general vector for the production of recombinant proteins. The vector introduces a *N*- or *C*-terminal $His_6$-tag upon in–frame ligation of the gene of interest, enabling Ni-NTA affinity chromatography based protein purification. Transcription of the cloned gene is carried out by the T7 RNA polymerase and is induced upon IPTG-mediated dissociation of the *lac*-repressor from the *lac*-operator sequence.

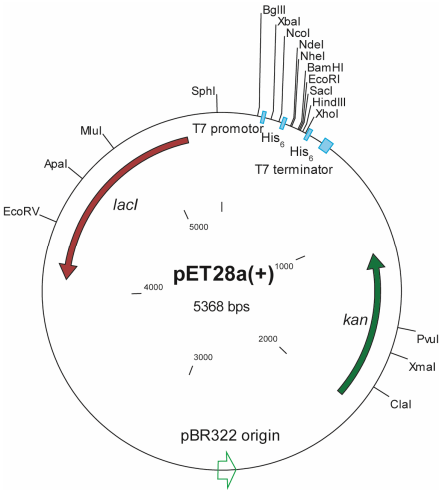



**Figure 3.2:** Physical map of pET28a(+).

### 3.3.3 pCB28a(+)

The pCB28a(+) vector is a derivatized pET28a(+) vector (Novagen) with a modified multiple cloning site. The vector appends a *N*- or *C*-terminal $His_6$-tag upon in-frame ligation of the targeted gene and allows purification of the recombinant protein via Ni-NTA affinity chromatography. The translation of the gene of interest is dependent on the T7 RNA polymerase and is induced upon addition of IPTG.

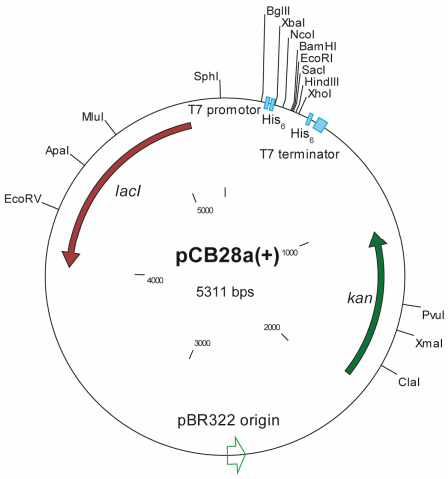



**Figure 3.3:** Physical map of pCB28a(+).

### 3.3.4 pREP4

The pREP4 plasmid (Stratagene) is commonly used in combination with the pQE expression system. It encodes for the *lac*-repressor, which binds to the *lac*-operator sequences on pQE vectors, giving rise to an IPTG-inducible expression system. The plasmid-borne *neo*-gene confers resistance to kanamycin up to a final concentration of 50 µg/mL.

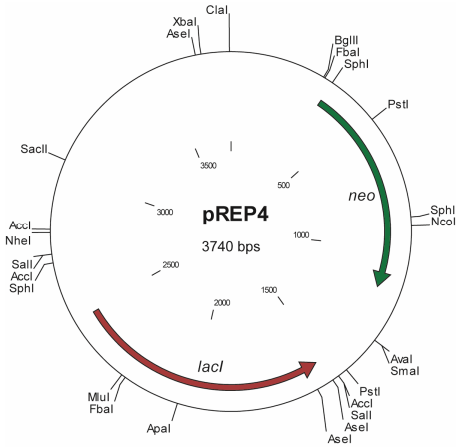



**Figure 3.4:** Physical map of pREP4.

### 3.4 Oligonucleotides

The oligonucleotides listed below were used for the PCR-based amplification of target genes. All primers were purchased from Sigma-Aldrich (*Steinheim, Germany*) or Eurofins MWG Operon (*Ebersberg, Germany*).

**Table 3.3:** Oligonucleotides utilized for the amplification of target genes. Restriction sites required for the subsequent construction of expression vectors are highlighted in red.

| oligonucleotide | sequence (5'→3') | restriction enzyme | vector | gene |
|---|---|---|---|---|
| 5'-*etcB* | GGAATTCCATATGACGCGGTATGACGACTCGAA | *Nde*I | pCB28a(+) | *etcB* |
| 3'-*etcB* | TTTTAAGCTTCTACTGCGCCACGCGCAGC | *Hind*III | | |
| 5'-*mcd* | TTTTTGCTAGCATGGCGGGCGACGTGGAACTC | *Nhe*I | pET28a(+) | *mcd* |
| 3'-*mcd* | TTTTTCTCGAGTCAGACCCGGGAGAAGACCATCAGA | *Xho*I | | |
| 5'-*sace_1309* | TTTTTGCTAGCATGCACGACATCATCGGCATCGGGT | *Nhe*I | pET28a(+) | *sace_1309* |
| 3'-*sace_1309* | TTTTTCTCGAGTCAGCTCCCGCTGAGCGCGTAG | *Xho*I | | |
| 5'-*etcC* | TTTTTGCTAGCGGGGGCGGCGCGTCCTCCGG | *Nhe*I | pET28a(+) | *etcC* |
| 3'-*etcC* | TTTTTCTCGAGCTACCGGGCCGGTCGCGCGAGCT | *Xho*I | | |
| 5'-*etcD PCP-C*$_4$ | TTTCTAGCTAGCGGCGACGACCGGCCGGCCACCGC | *Nhe*I | pET28a(+) | PCP-C$_4$ of *etcD* |
| 3'-*etcD PCP-C*$_4$ | TTTTCTCGAGTCACCAGGTCCTCCTAGGCGGATC | *Xho*I | | |

### 3.5 Microorganisms

#### 3.5.1 *E. coli* XL1-Blue

This genetically optimized strain (Stratagene) was used for cloning and sequencing purposes only. The genotype is as follows: *end*A1 *gyr*A96 *hsd*R17 *rec*A1 *rel*A1 *sup*E44 *thi*-1 *lac* F´[*pro*AB$^+$ *lacI*$^q$ *lac*ZΔM15 Tn*10*(Tet$^r$)].

#### 3.5.2 *E. coli* TOP10

*E. coli* TOP10 (QIAgen) was used as a general host for subcloning. The genetically optimized genotype is as follows: F$^-$ *mcr*A (*mrr-hsd*RMS-*mcr*BC) 80*lac*Z.M15*lac*X74 *deo*R *rec*A1 *ara*D139 (*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G.

### 3.5.3 *E. coli* M15/pREP4

*E. coli* M15/pREP4 (Invitrogen) was used as a host for the heterologous production of TioS PCP-TE. The lack of the Ion- and the membrane-associated OmpT-protease increases the half-life of heterologously produced proteins. The plasmid pREP4, which confers Kan-resistance, contains the Lac-repressor encoding gene *lacI*, enabling the controlled expression of pQE-constructs via IPTG-mediated induction. The genotype is given as follows: *nals strs rifs thi lac ara gal recA uvr mtl lon F-.*

### 3.5.4 *E. coli* BL21 (DE3)

*E. coli* BL21 (DE3) (Invitrogen) served as a general host for the expression of T7 promotor-based vectors. The lack of the Ion- and the OmpT-protease increases protein expression yields, due to reduced proteolytic degradation. The insertion of the IPTG-inducible T7 RNA polymerase after *lacZ* and the promoter *lac*UV5 on a λ-prophage enables T7 promotor driven expression. The genotype is as follows: $F^-$ *ompT gal dcm lon hsd*$S_B$($r_B^-$ $m_B^-$) λ(DE3 [*lacI lac*UV5-T7 gene 1 *ind*1 *sam*7 *nin*5]).

### 3.5.5 *Micromonospora sp*. L13-ACM2-092

*Micromonospora sp.* L13-ACM2-092 (CECT 3326) represents an aerobic Gram-positive, mesophilic actinomycete, which was isolated from a marine environment and was obtained from the Spanish Type Culture Collection (Colleción Española de Cultivos Tipo - CECT). The genome of the patented strain is not sequenced. The microorganism was cultivated for the isolation of the chromodepsipeptide thiocoraline and for the preparation of genomic DNA.

### 3.5.6 *Saccharopolyspora erythraea* NRRL 23338

The aerobic mesophilic Gram-positive filamentous actinomycete *Saccharopolyspora erythraea* NRRL 23338 is the natural producing strain of the macrolide polyketide antibiotic erythromycin and was obtained from the Agricultural Research Service (ARS). The recently annotated genome comprises 8.21 mega base pairs (Mbp) and contains at least 25 biosynthetic operons for the production of known or predicted secondary metabolites (GenBank accession number: AM420293). The strain was cultivated for the radio-LCMS-guided isolation of novel natural products and for the retrieval of genomic DNA.

### 3.6 Culture Media

The following media were used for the cultivation and fermentation of the microorganisms given above. For cultivation of the microorganisms on solid agar slants Agar No. 1 was added to a final concentration of 1.2% (w/v) to the corresponding medium. All media were autoclaved prior to inoculation (121°C, 1.5 bar, 20 min). Sterile glucose solution, starch, thiamin and biotin were added after autoclaving. Selection antibiotics were added to the autoclaved medium in the following concentrations: ampicillin 100 µg/mL, kanamycin 50 µg/mL or 25 µg/mL (*E. coli* M15/pREP4).

#### 3.6.1 LB Medium

All *E. coli* strains were grown in LB-medium (pH 7.5)

| **LB medium** | |
|---|---|
| **component** | **quantity (g/L)** |
| tryptone | 10 |
| yeast extract | 5 |
| sodium chloride | 5 |

#### 3.6.2 Medium 65

Medium 65 was used for the initial cultivation of the lyophilized *Micromonospora sp.* L13-ACM2-092 strain. The pH was adjusted to 7.2 with potassium hydroxide solution.

| **Medium 65** | |
|---|---|
| **component** | **quantity (g/L)** |
| malt extract | 10 |
| yeast extract | 4 |
| glucose | 4 |

### 3.6.3 *Micromonospora* Medium

This medium was used for the fermentation of *Micromonospora sp.* L13-ACM2-092 and for the subsequent extraction and purification of native thiocoraline.

| ***Micromonospora* medium** | |
|---|---|
| **component** | **quantity (g/L)** |
| malt extract | 10 |
| beef extract | 5 |
| yeast extract | 5 |
| glucose | 5 |
| starch | 20 |
| tryptone | 5 |
| sodium chloride | 4 |
| calcium carbonate | 4 |
| sodium sulfate | 1 |
| potassium chloride | 0.5 |
| dipotassium phosphate | 0.5 |
| magnesium chloride | 2 |

### 3.6.4 SCM Medium

The SCM medium was used for the cultivation of *Saccharopolyspora erythraea* NRRL 23338. Furthermore, radiolabeling experiments and supernatant extractions were carried out employing this medium.

| **SCM medium** | |
|---|---|
| **component** | **quantity (g/L)** |
| yeast extract | 1.5 |
| starch | 10 |
| soytone | 20 |
| MOPS | 10.5 |
| calcium chloride | 0.1 |

### 3.6.5 M9 Minimal Medium

The modified M9 minimal medium was used for the fermentative production and isolation of erythrochelin from cultures of *S. erythraea*. In addition, radiolabeling experiments were performed using this medium. Biotin (50 µg/L) and thiamin (200 µg/L) were added after autoclaving.

| **M9 minimal medium** | |
|---|---|
| **component** | **quantity (g/L)** |
| potassium dihydrogen phosphate | 3 |
| disodium phosphate | 6.78 |
| glucose | 2 |
| sodium chloride | 0.5 |
| ammonium chloride | 1.2 |
| magnesium sulfate | 0.12 |
| calcium chloride | 0.01 |
| glycerol | 0.1 |

# 4. Methods

## 4.1. Molecular Biology Methods

### 4.1.1 Cultivation of *Micromonospora sp.* L13-ACM2-092

The lyophilized cell culture of *Micromonospora sp.* L13-ACM2-092, as supplied by CECT, was incubated with 5 mL of Medium 65 (30 min, 25°C). Subsequently 200 mL of Medium 65 were inoculated with 250 µL of this solution. The cell cultures were grown for two days (28°C, 250 rpm) in baffled shaker flasks. Liquid cell cultures were utilized for the preparation of genomic DNA (see 4.1.3) or stocked as glycerol solutions (20% (v/v), -80°C). Medium-term cultivation of the actinomycete was carried out on Medium 65 solid agar slants.

### 4.1.2 Cultivation of *Saccharopolyspora erythraea* NRRL 23338

*Saccharopolyspora erythraea* NRRL 23338 spore stocks were obtained from the ARS culture collection and were initially incubated with 5 mL of SCM medium (30 min, 25°C). The rehydrated spores were transferred into 30 mL of SCM medium. The cultures were grown for five days (30°C, 250 rpm) in baffled shaker flasks. Liquid cell cultures were utilized for the preparation of genomic DNA (see 4.1.3) or stocked as glycerol solutions (20% (v/v), -80°C). Medium-term cultivation of the actinomycete was carried out on SCM solid agar slants.

### 4.1.3 Preparation of Genomic DNA

Liquid cell cultures of either *Micromonospora sp.* or *S. erythraea* (5 mL) were harvested by centrifugation (13,000 rpm, 3 min). Cell pellets were washed with water (1 mL) and resuspended in 500 µL lysis buffer (100 mM TRIS pH 8.0, 50 mM EDTA, 1% (w/v) SDS). Acid-washed glass beads were added to the suspension to obtain a final volume of 1.25 mL. The suspension was vortexed for 2 min and the liquid was separated from the glass beads. Subsequently, 275 µL of 7 M ammonium acetate solution (pH 7.0) were added and the solution was incubated for 5 min at 65°C. After 5 min incubation on ice, 500 µL of chloroform were added and the solution was shaken gently. Centrifugation (13,000 rpm, 5 min) allowed the transfer of the aqueous phase to a new reaction tube. The genomic DNA was precipitated by addition of 1 mL isopropyl alcohol and incubation at RT for 5 min. The DNA pellet obtained by centrifugation (13,000, 5 min) was washed twice (200 µL, 70% EtOH (v/v)), dried for 5 min at RT and solubilized in 200 µL $ddH_2O$. Genomic DNA was stored at 4°C.

### 4.1.4 Preparation of Plasmid DNA

4 mL of LB medium were inoculated with a single colony of the target strain and incubated over night (37°C, 250 rpm). 2 mL of cell culture were transferred into a 2 mL reaction tube and centrifuged (13,000 rpm, 3 min). The cell pellet was resuspended in 300 µL of buffer P1 (50 mM TRIS pH 8.0, 10 mM EDTA, RNase A 100 µg/mL). 300 µL of buffer P2 (0.2 M aq. NaOH, 1% (w/v) SDS) were added and the mixture was shaken gently and incubated at RT for 5 min. Subsequently, 300 µL of buffer P3 (2.55 M aq. KOAc, pH 4.8) were added and the reaction tube was incubated for 5 min on ice. Cell debris was removed form the mixture by centrifugation (13,000 rpm, 30 min) and the supernatant was transferred to a 1.5 mL reaction tube. 700 µL of cold isopropyl alcohol (4°C) were added and the plasmid DNA was precipitated by centrifugation (13,000, 4°C, 30 min). The supernatant was discarded and the DNA pellet was washed (200 µL, 70% EtOH (v/v)). Centrifugation (13,000 rpm, 4°C, 30 min) afforded the plasmid DNA pellet, which was dried (45°C, 15 min) and resolubilized in 40 µL of $ddH_2O$. Plasmid DNA was stored at -20°C.

### 4.1.5 Construction of Expression Plasmids

The amplification of target genes or gene fragments was carried out by polymerase chain reaction (PCR) with the Phusion Polymerase (New England Biolabs) with minor modifications to the manufacturers protocol. The DMSO concentration was 5% v/v and PCR reaction mixtures, all supplemented with GC buffer for amplification of GC-rich templates, were incubated at 98°C for 1 min prior to the addition of the polymerase. Purification of the PCR-products was performed with the QIAquick PCR purification kit (QIAgen). Digestion of the PCR-products and the corresponding plasmids with restriction endonucleases and ligation was carried out according to the manufacturers protocol (New England Biolabs). Ligations were transformed into *E. coli* TOP10 or XL-1 Blue cells via electroporation (2.5 kV). Transformants were plated on LB-agar slants supplemented with the corresponding antibiotic. The resulting plasmids were isolated as described above (4.1.4) or alternatively by utilization of the QIAprep Miniprep kit (QIAgen). Isolated restriction plasmids were verified by restriction mapping and dideoxy sequencing (GATC). Transformation of either *E. coli* M15/pREP4 with plasmids containing *tioS PCP-TE* or *E. coli* BL21 (DE3) with plasmids encoding *S. erythraea* genes was carried out according to the protocol supplied by the manufacturer (Invitrogen).

**Construction of pQE60(*tioS PCP-TE*)** The *tioS PCP-TE* gene fragment, encoding the PCP-TE didomain of TioS, was synthesized by EZBiolabs (*Westfield, USA*) and supplied as pBluescriptIISK(+)(*tioS PCP-TE*).

Codon usage was optimized for expression in *E. coli*. The gene was obtained by digestion of the vector with *Bam*HI and *Nco*I and cloned into the appropriate restriction sites of pQE60 (3.3.1).

**Construction of pCB28a(+)(*etcB*)** The gene *etcB*, encoding the FAD-dependent monooxygenase EtcB, was amplified from genomic DNA of *S. erythraea* by PCR employing the oligonucleotide primers listed in Table 3.3. The amplicon was digested with the corresponding restriction endonucleases and subsequently cloned into the appropriate restriction sites of pCB28a(+) (3.3.3).

**Construction of pET28a(+)(*mcd*)** The gene *mcd*, encoding the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd, was amplified from genomic DNA of *S. erythraea* by PCR employing the oligonucleotide primers listed in Table 3.3. The amplicon was digested with the corresponding restriction endonucleases and subsequently cloned into the appropriate restriction sites of pET28a(+) (3.3.2).

**Construction of pET28a(+)(*sace_1309*)** The gene *sace_1309*, encoding the FAD-dependent monooxygenase Sace_1309, was amplified from genomic DNA of *S. erythraea* by PCR employing the oligonucleotide primers listed in Table 3.3. The amplicon was digested with the corresponding restriction endonucleases and subsequently cloned into the appropriate restriction sites of pET28a(+) (3.3.2).

**Construction of pET28a(+)(*etcD PCP-$C_5$*)** The gene fragment *etcD PCP-$C_5$*, encoding the *C*-terminal PCP-C didomain of EtcD, was amplified from genomic DNA of *S. erythraea* by PCR employing the oligonucleotide primers listed in Table 3.3. The amplicon was digested with the corresponding restriction endonucleases and subsequently cloned into the appropriate restriction sites of pET28a(+) (3.3.2).

**Construction of pET28a(+)(*etcC)*** The gene fragment *etcC*, encoding the periplasmic binding protein EtcC, putatively mediating erythrochelin uptake, was amplified from genomic DNA of *S. erythraea* by PCR employing the oligonucleotide primers listed in Table 3.3. The amplicon was digested with the corresponding restriction endonucleases and subsequently cloned into the appropriate restriction sites of pET28a(+) (3.3.2).

## 4.2 Protein Purification and Analysis Methods

DNA-agarose gel analysis, SDS-PAGE and standard protein analysis methods were carried out according to established protocols.[197-198]

### 4.2.1 Protein Expression

#### 4.2.1.1 Expression of pQE60 Constructs

Expression of the pQE60(*tioS PCP-TE*) construct was performed by inoculating 500 mL of prewarmed (37°C) LB medium, supplemented with ampicillin (100 µg/mL) and kanamycin (25 µg/mL) in a 2 L baffled shaker flask with 5 mL of an overnight culture of the *E. coli* M15/pREP4. The cells were grown to an optical density (OD) of 0.6 ($\lambda$ = 600 nm) at 37°C and 250 rpm, induced with 0.1 mM isopropyl-$\beta$-D-thiogalactopyranoside (IPTG) and cultivated for an additional 5 h at 20°C. The cells were harvested by centrifugation (6000 rpm, 4°C, 30 min), resuspended in 40 mL buffer (50 mM HEPES, 250 mM NaCl, pH 7.5) and either directly processed or stored at -20°C.

#### 4.2.1.2 Expression of pET28a(+) and pCB28a(+) Constructs

Expression of the pET28a(+) or pCB28a(+) constructs was carried out by inoculating 500 mL of prewarmed (37°C) LB medium, supplemented with kanamycin (50 µg/mL) in a 2 L baffled shaker flask with 5 mL of an overnight culture of the *E. coli* BL21 (DE3). The cells were grown to an optical density (OD) of 0.6 ($\lambda$ = 600 nm) at 37°C and 250 rpm, induced with 0.1 mM IPTG and cultivated for ~ 18 h at 20°C. The cells were harvested by centrifugation (6000 rpm, 4°C, 30 min), resuspended in 40 mL buffer (50 mM HEPES, 250 mM NaCl, pH 7.5) and either directly processed or stored at -20°C.

### 4.2.2 Protein Purification

The purification of recombinant $His_6$-tagged proteins was initiated by disruption of the cells using either an EmulsiFlex-C5 High Pressure Homogenizer (Avestin) or a French Pressure Cell (SLM Aminco). Cell debris and insoluble components were precipitated by centrifugation (17,000 rpm, 4°C, 30 min), the supernatant conztaining the protein was separated and the protein was purified by Ni-NTA (QIAgen) affinity chromatography using either a FPLC system 250 (Amersham Pharmacia Biotech) or an Äkta prime system (Amersham Pharmacia Biotech). The column was equilibrated with buffer A (50 mM HEPES, 250 mM NaCl, pH 7.5). The supernatant was applied onto the column with a flow-rate of 0.7 mL/min until the absorption of the flowthrough at 280 nm ($A_{280}$) reached baseline level. Selective elution of the

immobilized protein was performed by application of following solvent gradient of buffer A and buffer B (50 mM HEPES, 250 mM NaCl, 250 mM imidazole, pH 7.5) at a flow rate of 1 mL/min: linear increase from 0% to 50% B within 35 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. The wavelength chosen for detection of the protein was 280 nm and the fraction size was 2 mL. Protein-containing fractions were identified via qualitative Bradford assays and analyzed by SDS-PAGE.[197,199] Fractions containing the recombinant protein were combined, concentrated with Amicon Ultra-15 concentrators (Millipore) and subjected to buffer exchange utilizing HiTrap Desalting columns (GE Healthcare). Buffers for recombinant EtcB and Sace_1309 were supplemented with glycerol (20% (v/v), pH 8.0). Protein solutions were subsequently aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.

### 4.2.3 Size-Exclusion Chromatography (SEC)

Size-exclusion chromatography (SEC) was employed as a second purification step to obtain the periplasmic binding protein EtcC in high purity. Gelfiltration enables the fractionation of proteins based on their relative size, which directly correlates with the diffusion behaviour in a gel-matrix. The column used was a 16/60 Superdex G75 prep grade (Amersham Biosciences) gelfiltration column and was equilibrated with two column volumes of buffer (25 mM Hepes, 50 mM NaCl, pH 7.0). Subsequently, the protein solution was applied onto the column at a flow rate of 1 mL/min. Elution was performed with a flow rate of 1 mL/min and the wavelength chosen for detection of the protein was 280 nm. The fraction size was 2 mL and protein containing fractions were identified via qualitative Bradford assays and analyzed by SDS-PAGE.[197,199] Fractions containing the recombinant protein were combined, concentrated with Amicon Ultra-15 concentrators (Millipore), aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.

### 4.2.4 Protein Concentration Determination

Protein concentrations were determined spectrophotometrically using calculated extinction coefficients after thawing of the corresponding protein solution. The coefficients were calculated by analysis of the primary sequence with Protean (DNAstar). The determination was carried out on a Nanodrop ND-1000 (PEQLab) spectrophotometer ($\lambda$ = 280 nm).

**Table 4.1:** The calculated molar extinction coefficients (λ = 280 nm). The corresponding mass is given for each protein.

| protein | calculated extinction coefficient ($M^{-1}$ $cm^{-1}$ $10^3$) | MW (kDa) |
|---|---|---|
| TioS PCP-TE | 46.5 | 38.4 |
| EtcB | 41.8 | 51.3 |
| Mcd | 64.3 | 48.4 |
| Sace_1309 | 33.9 | 40.9 |
| EtcC | 46.4 | 36.1 |
| EtcD PCP-$C_5$ | 52.0 | 58.3 |

### 4.3 Analytical Methods

#### 4.3.1 MALDI-TOF-MS

Matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS) represents a mass spectrometric method for the determination of the molecular mass of ionized compounds in high vacuum. Ionization of the analyte is achieved by excitation of a matrix in which the analyte is embedded with distinct laser pulses. The energy of the laser is gently transferred to the analyte, circumventing fragmentation processes and enabling analysis of high molecular weight samples. The mass spectrometer used is most commonly a TOF (time of flight) spectrometer equipped with an ion mirror for increased resolution. Samples were prepared by mixing 0.5 µL of analyte with 0.5 µL of DHB matrix solution (Agilent Technologies). The mixture was subsequently transferred onto a metallic target and incubated at 37°C for 15 min. After evaporation of the solvent the sample was measured on a Bruker BiFlex III MALDI-TOF system.

#### 4.3.2 RP-HPLC and LCMS

Liquid chromatography mass spectrometry (LCMS) was used as a standard method for the characterization of substrates and products based on retention time on a chromatography column and by mass spectrometric analysis. Reversed-phase (RP) columns were routinely employed and retained the compounds due to hydrophobic interactions of the analyte with the non-polar stationary phase. Stationary phases consisted of carbon or of hydrophobic alkyl chains ($C_{18}$, $C_8$ or $C_4$) immobilized on silica gel. Residual silanols were endcapped by the modification with trimethylchlorosilane (TMS) in the case of ec-columns. Elution of the compounds was performed with acetonitrile or methanol, which compete with the analyte for non-covalent interaction with the column material. The mobile phases were supplemented with 0.1% TFA (trifluoroacetic acid), 0.05% formic acid or 20 mM nonafluoropentanoic

acid (NFPA) as ion pair reagents and enabled ionization. The retention time of the compound was monitored by UV/Vis- or fluorescence-detection. The electrospray ionization mass detector allowed ionization and mass detection at atmospheric pressure. General experiments were carried out on a Agilent 1100 system. Peptide mass fingerprinting analysis was carried out on a quadrupole time-of-flight (QTOF) instrument.

#### 4.3.3 HRMS and MS/MS-Fragmentation Analysis

High resolution mass spectrometric characterization and MS/MS-fragmentation analysis of products of enzymatic reactions, erythrochelin and *ferri*-erythrochelin was performed with an LTQ-FT instrument (Thermo Fisher Scientific) connected to a microbore Agilent 1100 HPLC system. The compounds were analyzed on a Nucleodur $C_{18}$(ec) 125 x 2 mm column utilizing the following solvent gradient: 0-30 min, 0%-100% acetonitrile into water, both supplemented with 0.1% TFA. The column temperature was 45°C and the flow-rate was 0.3 mL $min^{-1}$. CID fragmentation studies within the linear ion trap were done using online LCMS.

#### 4.3.4 Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) was utilized to validate the identity of the isolated recombinant proteins. Gel bands of the correct size were excised from the gel after SDS-PAGE and incubated with wash solution (200 µL, 200 mM $NH_4HCO_3$, 50% v/v MeCN) for 30 min at 37°C. The solvent was removed to complete dryness in a speed-vac manifold (37°C, 30 min). In-band tryptic digestion was carried out by addition of trypsin solution (20 µL, 0.02 µg/µL trypsin, 10% $NH_4HCO_3$, 10% (v/v) MeCN, pH 8.1) to the dried gel bands. Excess trypsin solution was removed after initial incubation at 37°C for 45 min, followed by incubation for an additional 16 h. Peptide fragments derived from proteolytic cleavage were eluted with diffusion solution (25 µL, 1% (v/v) TFA, 10% (v/v) MeCN, pH 8.1) in parallel to sonication (45 min, RT). The samples were analyzed by the mass spectrometry facility (Department of Chemistry, Philipps-University Marburg) on a nano spray-HPLC-QTOF-MS system. Peptide mass fingerprint comparison with the MASCOT database enabled protein identification.

### 4.4 Spectroscopic Methods

#### 4.4.1 UV/Vis-Spectroscopy

UV/Vis-spectra were recorded on an Ultrospec 3000 (Pharmacia) spectrophotometer. Wavescan measurements were performed within a wavelength range of 200-800 nm and a scan rate of

750 nm/min. Absorption spectra of erythrochelin and *ferri*-erythrochelin were recorded in water. *Ferri*-erythrochelin was obtained by incubating erythrochelin (400 µM) with increasing amounts of aqeous $FeCl_3$ for 10 min at RT prior to the scan. Absorption spectra of EtcB and Sace_1309 (50 µM) were recorded to analyze the cofactor content of the recombinant proteins.

### 4.4.2 CD-Spectroscopy

CD spectra were recorded to investigate the native fold of recombinant TioS PCP-TE. The recombinant thioesterase GrsB TE served as a control for the comparison of the spectra. CD spectra were recorded on a J-810 spectropolarimeter (Jasco) with a final concentration of 5 µM TioS PCP-TE or GrsB TE at 20 °C with 20 nm/min, response of 2 s and a bandwidth of 1 nm in 10 mM $Na_2HPO_4$ buffer (pH 7.0). Far-UV-CD spectra were recorded in a wavelength range of 180-280 nm. The data presented is an accumulation of ten scanning cycles. Mean residue weight ellipticity ($[\Theta]_{MRW}$) was calculated from the measured ellipticity ($\Theta$) according to the following equation (1-1).

$$[\Theta]_{MRW} = \frac{\Theta}{c \cdot d \cdot N_{AA} \cdot 10} \qquad (1\text{-}1)$$

$[\Theta]_{MRW}$ mean residue weight ellipticity (deg $cm^2$ $dmol^{-1}$)
$\Theta$ measured ellipticity (mdeg)
c protein concentration (M)
d light path (cm)
$N_{AA}$ amino acid residues

### 4.4.3 NMR-Spectroscopy and Structure Elucidation

NMR-spectroscopic structure elucidation of erythrochelin was performed in cooperation with Dr. Xiulan Xie (Department of Chemistry, Philipps-University Marburg). All spectra were recorded on an AV600-spectrometer (Bruker) to elucidate the structure of erythrochelin. 16 mg of the title compound was dissolved in 0.7 ml DMSO-$d_6$. Measurements were carried out with an inverse broadband probe installed with z gradient. The one-dimensional spectra $^1H$ and $^{13}C$, the homonuclear two-dimensional spectra DQF-COSY, TOCSY, NOESY, and ROESY, the $^1H$– $^{13}C$ HSQC and HMBC, and the $^1H$–$^{15}N$ HSQC spectra were recorded at room temperature with standard pulse programs.[200] Phase-sensitive HMBC spectrum focused on the carbonyl region with high resolution in $^{13}C$ dimension was recorded by using a pulse program with a semi-selective $^{13}C$ pulse built in an HMBC experiment with sensitivity enhancement.[201-202] The TOCSY spectrum was recorded with a mixing time of 200 ms, while NOESY and ROESY spectra were taken at 150 and 300 ms mixing times. The one- dimensional spectra were acquired with 65536 data

points, while two-dimensional spectra were collected using 4096 points in the $F_2$ dimension and 512 increments in the $F_1$ dimension. For two-dimensional spectra 16 - 32 transients were used. Relaxation delay was 2.5 s. Chemical shifts of $^{1}H$ and $^{13}C$ were referenced to the solvent signals, while that of $^{15}N$ was referenced externally to the urea signal. All recorded spectra were processed with Bruker Topspin 2.1.

## 4.5 Chemical Synthesis

### 4.5.1 Solid-Phase Peptide Synthesis of Tetrapeptidyl Substrates

Tetrapeptidyl substrates employed in macrocyclization assays were obtained by solid-phase peptide synthesis (SPPS) on an APEX 396 synthesizer (Advanced ChemTech) (0.1 mmol scale) with 2-chlorotrityl resin as solid support, employing Fmoc-protective group strategy (Figure 4.1). The reaction chamber was vortexed after each reaction step and remaining solvents were removed by applying a pressure of 62 kPa. In the first step of a typical SPPS-cycle the amino acid was loaded onto the solid support and the Fmoc-group, protecting the $\alpha$-$NH_2$ group, was cleaved with piperidine. Subsequently, elongation was initiated by addition of DIPEA, the coupling reagent HBTU, the additive HOBt and the *N*-terminally protected amino acid. The assembled oligopeptide was cleaved from the resin by incubation with cleavage solution (AcOH:TFE:DCM, 1:2:7 (v/v)). The acidic components of the cleavage solution were removed azeotropically with hexane and the peptide was used without further purification for the preparation of *C*-terminally activated macrocyclization substrates.

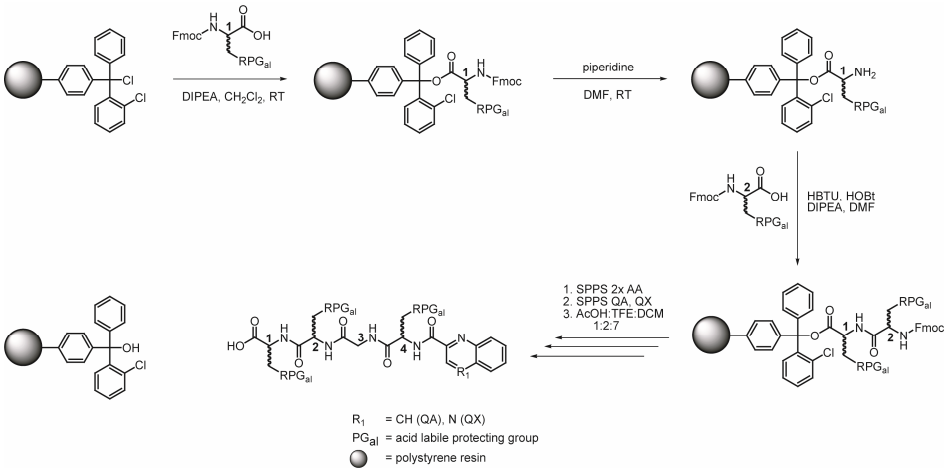



**Figure 4.1**: Solid-phase peptide synthesis (SPPS) of the linear tetrapeptidyl substrate. The synthesis is initiated by the immobilization of the first amino acid on the solid support. Consecutive deprotection and coupling reactions result in the resin-bound tetrapeptide. Prior to the cleavage of the assembled peptide, a last SPPS cycle caps the *N*-terminus of the peptide with the chromophore moiety QA or QX.

### 4.5.2 Synthesis of Tetrapeptidyl-SNAC Substrates

*C*-terminal activation of the tetrapeptidyl substrates as SNAC-thioesters was carried out chemically on a 50 µmol-scale. The protected peptide (0.05 mmol, 1.0 eq.) was dissolved in dichloromethane (4 mL) and stirred with PyBOP (31.2 mg, 0.06 mmol, 1.2 eq.) and *N*-acetylcysteamine (14.3 mg, 0.12 mmol, 2.4 eq.) at RT for 5 min. Subsequently, DIPEA (10.4 µL, 0.06 mmol, 1.2 eq.) was added and the solution was stirred for 2 h (Figure 4.2). Side-chain deprotection (Boc, Trt) was carried out by addition of cleavage solution (4 mL, TFA:TIPS:dd$H_2O$, 95:2.5:2.5 (v/v)) and stirring for an additional 2 h. The reaction mixture was poured into diethylether (42 mL, -20°C) and the peptide precipitated over night at -20°C. The solvent was removed after centrifugation (4000 rpm, 4°C, 10 min). The crude product was dissolved in 10% MeCN and subsequently applied onto a RP-HPLC preparative Nucleodur $C_{18}$(ec) 250 × 21 mm column (Macherey & Nagel) combined with an Agilent HPLC-system. Elution was performed by application of the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) at a flow rate of 16 mL/min: linear increase from 10% to 95% B within 45 min, holding B for an additional 5 min. Retention times were monitored at 215 nm and 280 nm. Product identity was confirmed by MALDI-TOF-MS and LCMS (Table 4.2). Product containing fractions were combined and lyophilized. Peptides were solubilized in dimethylsulfoxide (DMSO) to a final concentration of 20 mM and stored at -20°C.

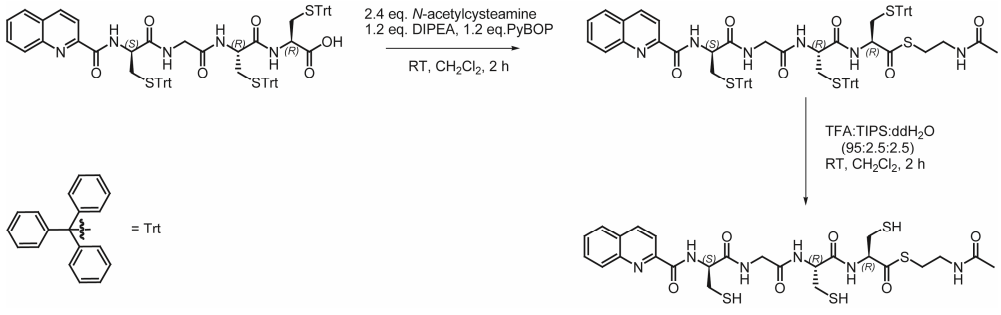



**Figure 4.2:** Synthesis of the activated tetrapeptidyl-substrates exemplified by activation and deprotection of substrate TL1: In the first step of the reaction the carboxyl group is activated by the coupling reagent PyBOP and thioesterified with *N*-acetylcysteamin. Removal of acid labile side-chain protecting groups affords the activated and deprotected substrate for macrocyclization studies.

**Table 4.2:** ESI-MS identification of the and *C*-terminally activated tetrapeptidyl thioesters.

| compound | sequence | species (m/z) | observed mass (calculated mass) |
|---|---|---|---|
| TL1 | QA-D-$Cys_1$-$Gly_2$-L-$Cys_3$-L-$Cys_4$-SNAC | $[M+H]^+$ | 641.1 (641.1) |
| TL2 | QA-D-$Cys_1$-$Gly_2$-L-$Cys_3$-*S*-Me-L-$Cys_4$-SNAC | $[M+H]^+$ | 655.2 (655.1) |
| TL3 | QA-D-$Cys_1$-$Gly_2$-L-$Ala_3$-*S*-Me-L-$Cys_4$-SNAC | $[M+H]^+$ | 623.2 (623.2) |
| TL4 | QA-D-$Cys_1$-$Gly_2$-L-$Ala_3$-L-$Met_4$-SNAC | $[M+H]^+$ | 637.3 (637.2) |
| TL5 | QA-D-$Cys_1$-$Gly_2$-L-$Cys_3$-L-$Met_4$-SNAC | $[M+H]^+$ | 669.3 (669.2) |
| TL6 | QA-L-$Cys_1$-$Gly_2$-L-$Ala_3$-L-$Met_4$-SNAC | $[M+H]^+$ | 637.2 (637.2) |
| TL7 | QA-D-$Ser_1$-$Gly_2$-L-$Cys_3$-L-$Cys_4$-SNAC | $[M+H]^+$ | 625.1 (625.2) |
| TL8 | QA-D-$Ser_1$-$Gly_2$-L-$Cys_3$-*S*-Me-L-$Cys_4$-SNAC | $[M+H]^+$ | 639.3 (639.2) |
| TL9 | QA-D-$Ser_1$-$Gly_2$-L-$Ala_3$-*S*-Me-L-$Cys_4$-SNAC | $[M+H]^+$ | 607.3 (607.2) |
| TL10 | QA-D-$Ser_1$-$Gly_2$-L-$Ala_3$-L-$Met_4$-SNAC | $[M+H]^+$ | 621.3 (621.2) |
| TL11 | QX-D-$Cys_1$-$Gly_2$-L-$Ala_3$-*S*-Me-L-$Cys_4$-SNAC | $[M+H]^+$ | 624.3 (624.2) |

## 4.6 Biochemical Methods

### 4.6.1 Fluoresceinyl-CoA Phosphopantetheinylation Assay

The PCP-domain is converted from inactive *apo*- to active *holo*-PCP by transfer of the ppan group onto the core-T motif serine, which is mediated by PPtases. The extensively characterized PPtase Sfp (*Bacillus subtilis*) displays a high degree of substrate tolerance and can catalyze the transfer of CoA-derivatized substrates onto the PCP *in vitro*.[203] The fluoresceinyl-CoA phoshopantetheinylation assay was carried out to ensure the possibility to posttranslationally load synthetic CoA-substrates onto the PCP-domain of EtcD PCP-$C_5$ (Figure 4.3).

Fluorescence labeling was achieved by incubating the recombinant protein EtcD PCP-$C_4$ (50 µM) with fluoresceinyl-CoA (300 µM), Sfp (5 µM) and $Mg^{2+}$ (50 µM) in buffer (20 mM TRIS, 100 mM NaCl, pH 7.0) for 30 min at 25°C. The control lacked the PPtase Sfp. Labeled protein was visualized on a UV-screen (λ = 312 nm) after SDS-PAGE. The gel was subsequently stained with Coomassie Brilliant Blue R250 for comparison of the labeled and unlabeled samples.

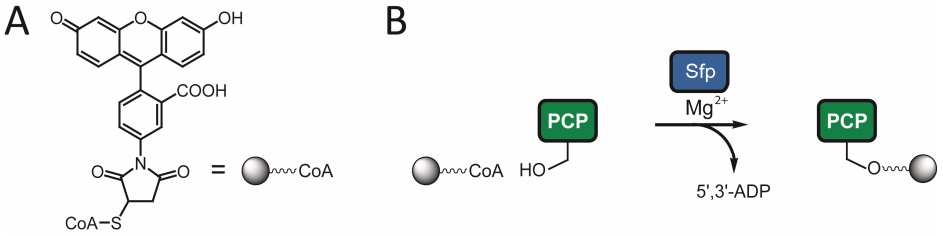



**Figure 4.3:** A) Chemical structure of fluoresceinyl-CoA. B) Sfp-mediated labeling of recombinant PCP with fluoresceinyl-CoA: The fluorescent probe is transferred onto the conserved core-T motif serine in the presence of $Mg^{2+}$.

### 4.6.2 Activity-Based Fluorescence Labeling Assay

Activity-based protein profiling represents an approach for the identification and isolation of enzymes from a complex proteome based on their specific activity. The serine protease superfamily was the first enzyme class to be analyzed employing this method. The activity-based probe (ABP) of choice was FP-TAMRA (fluorophosphonate-*N,N,N´,N´*-tetramethylrhodamine) which reacts with the active-site serine of the catalytic triad consisting of Ser-His-Asp.[204] The nucleophilic oxyanion of the serine side chain attacks the fluorophosphonate group under the release of fluoride, leading to covalent labelling of the target protein (Figure 4.4).

The activity-based fluorescence labeling assay was carried out to confirm the functionality of the catalytic triads of TioS PCP-TE, GrsB PCP-TE and GrsB TE. Covalent labeling was achieved by incubating the corresponding thioesterase (10 µM) with FP-TAMRA (30 µM) in buffer (25 mM NaCl, 50 mM HEPES, pH 7.0) in a total volume of 50 µL for 45 min at 25°C. Control samples were denatured at 95°C for 15 min prior to labelling. Labeled proteins were visualized on a UV-screen (λ = 312 nm) after SDS-PAGE. The gels were subsequently stained with Coomassie Brilliant Blue R250 for comparison of the labeled and unlabeled samples.

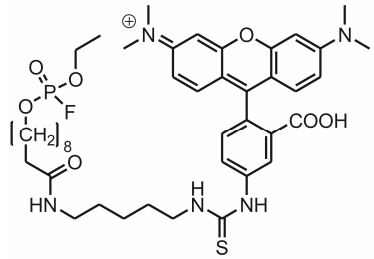



**Figure 4.4:** The chemical structure of the activity-based probe FP-TAMRA. The active-site serine of a thioesterase reacts with the fluorophosphonate group under the release of fluoride, resulting in a covalently labeled serine protease.

### 4.6.3 TioS PCP-TE-Catalyzed Macrocyclization Assay

The thioesterase TioS PCP-TE catalyzes the cyclodimerization of *C*-terminally SNAC activated tetrapeptidyl substrates. To investigate the macrocyclization potential of TioS PCP-TE, the recombinant enzyme was incubated with corresponding substrates.

Macrocyclization assays were carried out in a total volume of 50 µL in assay buffer (25 mM HEPES, 50 mM NaCl, pH 6.0) at 25°C. For temperature dependence evaluation of macrocycle formation the temperature was altered to 15°C or 37°C, respectively. Final concentration of substrate was 300 µM and total concentration of DMSO was 8% (v/v). The assays were initiated by the addition of 10 µM TioS PCP-

TE and quenched after 2 h by the addition of 10 µL 4% (v/v) TFA. Reduction of oxidatively formed disulfide-bonds was accomplished by the addition of 300 µM tris-(carboxyethyl)-phosphine (TCEP) in DMSO. Assays were analyzed by RP-LCMS (Agilent) on a Nucleodur $C_{18}$(ec) 125 × 2 mm column (Macherey & Nagel) utilizing the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) at a flow rate of 0.2 mL/min: linear increase from 20% to 80% B within 50 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. The wavelength chosen for detection was 210 nm and the column temperature was 45°C. Identities of the products were confirmed by LCMS. Kinetics of total substrate turnover were performed by determining initial rates of nine substrate concentrations, using three time points at each concentration within the linear region of the reaction. The concentration of linear tetrapeptidyl-thioester substrates was calculated on the basis of experimentally determined absorption values at a wavelength of 210 nm. Assays were analyzed by RP-LCMS as stated above and the kinetic parameters were determined with the Enzyme Kinetics Module for SigmaPlot 8.0 (Systat Software Inc.).

#### 4.6.4 Monooxygenase-Mediated Hydroxylation Assay

The FAD-dependent monooxygenases EtcB and Sace_1309 catalyze the hydroxylation of the δ-amino functionality of L-ornithine in presence of the cosubstrate NADPH.

Hydroxylation assays were carried out by incubating recombinant EtcB or Sace_1309 (20 µM) with L-ornithine (1 mM), the cosubstrate NADPH (1 mM) and the cofactor FAD (20 µM) in assay buffer (20 mM TRIS, 100 mM NaCl, pH 8.0) for 4-8 h at 25°C. The reactions were stopped by addition of 10 µL 4% (v/v) TFA. Hydroxylation assays were analyzed by RP-HPLC (Agilent) on a Hypercarb 100 x 2.1 mm column (Thermo) utilizing the following solvent gradient of solvent A (20 mM NFPA) and solvent B (acetonitrile) at a flow rate of 0.4 mL/min: linear increase from 0% to 15% B within 20 min followed by a linear increase to 100% B in 2 min, holding B for an additional 5 min. The wavelength chosen for detection was 215 nm and the column temperature was 20°C. Substrate specificity assays were performed by employing the same conditions stated above with different amino acids as substrates and an incubation time of 8 h. Specificity assays were analyzed by RP-LCMS (Agilent) on a Nucleodur $C_{18}$(ec) 125 × 2 mm column (Macherey & Nagel) utilizing the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) at a flow rate of 0.2 mL/min: linear increase from 0% to 100% B within 30 min holding B for an additional 5 min. The wavelength chosen for detection was 210 nm and the column temperature was 45°C. Reaction kinetics were determined spectrophotometrically on an Ultrospec 3100 pro spectrophotometer (Amersham Biosciences). Concentration of L-Orn was

varied between 75 µM and 20 mM. Calculation of initial reaction rates was based on the decrease of NADPH absorbance ($\varepsilon$ = 6300 $M^{-1}$ $cm^{-1}$) and the determination of kinetic parameters was carried out with the Enzyme Kinetics Module for SigmaPlot 8.0 (Systat Software Inc.).

### 4.6.5 Mcd-Mediated Acetylation Assay

Acetylation of the δ-amino functionality of L-hOrn is mediated by the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd encoded in the *nrps1* gene cluster, as shown via *in vitro*-acetylation assays.

Acetylation assays were carried out by incubating recombinant Mcd (50 µM) with δ-*N*-hydroxy-L-ornithine (1 mM) and the cosubstrates malonyl-CoA (1 mM) or acetyl-CoA (1 mM) in assay buffer (20 mM TRIS, 100 mM NaCl, pH 7.0) for 4-8 h at 25°C. The reaction was stopped by addition of 10 µL 4% (v/v) TFA. Acetylation assays were analyzed by RP-HPLC (Agilent) on a Hypercarb 100 x 2.1 mm column (Thermo) utilizing the following solvent gradient of solvent A (20 mM NFPA) and solvent B (acetonitrile) at a flow rate of 0.4 mL/min: linear increase from 0% to 15% B within 20 min followed by a linear increase to 100% B in 2 min, holding B for an additional 5 min. The wavelength chosen for detection was 215 nm and the column temperature was 20°C. Substrate specificity assays were performed by employing the same conditions stated above with different amino acids as substrates and an incubation time of 8 h. Specificity assays were analyzed by RP-LCMS (Agilent) on a Nucleodur $C_{18}$(ec) 125 × 2 mm column (Macherey & Nagel) utilizing the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) at a flow rate of 0.2 mL/min: linear increase from 0% to 100% B within 30 min holding B for an additional 5 min. The wavelength chosen for detection was 210 nm and the column temperature 45°C. Time-course experiments to monitor the conversion rate of L-hOrn to L-haOrn was accomplished by incubating recombinant Mcd (1 µM) with L-hOrn (1 mM) and the cosubstrates malonyl-CoA (1 mM) or acetyl-CoA (1 mM) in assay buffer (20 mM TRIS, 100 mM NaCl, pH 7.0) for 4 h at 25°C. Samples were taken each 30 min and analyzed via LCMS as described above.

### 4.6.6 Coupled Hydroxylation and Acetylation Assay

Coupled hydroxylation and acetylation assays were carried out to establish a biosynthetic route for the generation of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine.

Coupled assays were conducted by incubating recombinant EtcB or Sace_1309 (20 µM), Mcd (50 µM), L-ornithine (1 mM), the cosubstrates NADPH (1 mM), malonyl-CoA (1 mM) or acetyl-CoA (1 mM) and the

cofactor FAD (20 µM) in assay buffer (20 mM TRIS, 100 mM NaCl, pH 7.0) for 4-8 h at 25°C. The reaction was quenched by addition of 10 µL 4% (v/v) TFA. Product identities were analyzed by RP-HPLC (Agilent) on a Hypercarb 100 x 2.1 mm column (Thermo) utilizing the following solvent gradient of solvent A (20 mM NFPA) and solvent B (acetonitrile) at a flow rate of 0.4 mL/min: linear increase from 0% to 15% B within 20 min followed by a linear increase to 100% B in 2 min, holding B for an additional 5 min. The wavelength chosen for detection was 215 nm and the column temperature was 20°C.

### 4.7 Radio-LCMS-Guided Genome Mining

In an attempt to substitute the traditional activity assay-guided isolation of novel secondary metabolites a dedicated radio-LCMS-methodology to identify NRPs of cryptic gene clusters in cultures of S. *erythraea* was established. Radiolabeling studies were performed by cultivating *S. erythraea* in 100 mL of SCM medium (Chapter 3.6.4) or iron-deficient M9 medium (Chapter 3.6.5). 5 µCi of L-ornithine (Hartmann Analytic) were added after 48 h of growth. The supernatants were extracted with XAD16 resin after two additional days of growth. The dried eluate was dissolved in 10% methanol and analyzed on a Nucleodur $C_{18}$(ec) column 125 x 2 mm (Macherey & Nagel) combined with an Agilent 1100 HPLC-system (Agilent) connected to a FlowStar LB513 radioactivity flow-through detector (Berthold) equipped with a YG-40-U5M solid microbore cell and a QStar Pulsar i (Applied Biosystems) utilizing the following solvent gradient: water/0.05% formic acid (solvent A) and methanol/0.05% formic acid (solvent B) at a flow rate of 0.3 mL/min: linear increase from 0% to 50% B within 20 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. This gradient was also used to analyze comparative extractions of *S. erythraea* cultures and erythrochelin and *ferri*-erythrochelin.

### 4.8 Natural Product Isolation

#### 4.8.1 Isolation of Chemoenzymatically Generated Thiocoraline Analogs

For the isolation of macrocycles produced *in vitro*, the corresponding tetrapeptidyl-thioesters were incubated with the recombinant protein under standard assay conditions (Chapter 4.6.3) except for the formation of Cy8SS lacking TCEP. For maximum cyclization yields the experimentally determined optimal temperature was chosen. Enzymatic assays were analyzed and separated by RP-HPLC (Agilent) and product fractions were collected using a AnalytFC fraction collector (Agilent). Product yield and identity were determined by RP-HPLC (Agilent) on a Nucleodur $C_{18}$(ec) 125 × 2 mm column (Macherey & Nagel) utilizing the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA

(solvent B) at a flow rate of 0.2 mL/min: linear increase from 20% to 80% B within 50 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. The wavelength chosen for detection was 210 nm and the column temperature 45°C. Product fractions were lyophilized (Alpha 2-4 GSC, Christ), resolubilized in DMSO to a final concentration of 5 mM and stored at -20°C.

### 4.8.2 Isolation of Native Thiocoraline from *Micromonospora sp.*

*Micromonospora* L13-ACM2-092 (Chapter 3.5.5) was cultivated in *Micromonospora* medium (Chapter 3.6.3) under the conditions described earlier.[142] Cells were harvested by centrifugation and the mycelial cake and the supernatant were extracted with ethylacetate (EtOAc) twice. The combined organic layers were dessicated over $MgSO_4$ and evaporated to dryness. The crude extract was dissolved in 20% MeCN and subsequently applied onto a RP-HPLC preparative Nucleodur $C_{18}$(ec) 250 × 21 mm column (Macherey & Nagel) combined with an Agilent HPLC-system. Elution was performed by application of the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) at a flow rate of 16 mL/min: linear increase from 20% to 80% B within 50 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. Product identity was confirmed by LCMS and fluorescence detection (Agilent 1100 FLD; excitation 365 nm, emission 540 nm) according to established protocols. The retention time of thiocoraline was 42.3 min. Total yield was 1.6 mg thiocoraline/L culture.

### 4.8.3 Isolation of Erythrochelin from SCM Medium

*Saccharopolyspora erythraea* NRRL 23338 (3.5.6) maintained on SCM-agar slants, was used to inoculate 30 mL of SCM liquid culture (3.6.4). The cells were grown for 4 d at 30°C and 250 rpm and subsequently used to inoculate 1 L of SCM-medium. The cells were grown for 5 d at 30°C. The production phase of the strain was monitored via LCMS and the CAS assay.[205] The culture supernatant was extracted with XAD16 resin (4.0 g/L). The resin was collected by filtration, washed twice with water and the absorbed compounds were eluted with methanol. The eluate was evaporated to dryness, dissolved in 10% acetonitrile and applied onto a RP-HPLC preparative Nucleodur $C_{18}$(ec) 250 × 21 mm column combined with an Agilent 1100 HPLC-system. Elution was performed by application of the following solvent gradient of water/0.05% formic acid (solvent A) and methanol/0.05% formic acid (solvent B) at a flow rate of 16 mL/min: linear increase from 0% to 50% B within 50 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. The wavelengths chosen for detection were 215 nm and

280 nm, respectively. Siderophore containing fractions were confirmed by using the CAS-liquid assay and subjected to LCMS-analysis.

#### 4.8.4 Large-Scale Purification of Erythrochelin from M9 Medium

*S. erythraea*, maintained on SCM-agar slants, was used to inoculate 30 mL of SCM liquid culture (Chapter 3.6.4). The cells were grown for 4 d at 30°C and 250 rpm. The cells were exchanged from the SCM-medium into the iron-deficient M9 medium (Chapter 3.6.5) by repeated centrifugation and resuspension of the cells in the target medium. Subsequently, the cells were used to inoculate 5 L of iron-deficient M9 medium in PC-flasks to a final optical density ($\lambda$ = 600 nm) of 0.01. After 4 d of cultivation the cells were harvested by centrifugation at 6000 rpm and 4°C for 30 min. The supernatant was separated from the cell pellet and incubated with XAD16 resin (4.0 g/L). The resin was collected by filtration, washed twice with water and the absorbed compounds were eluted with methanol. The eluate was evaporated to dryness, dissolved in 10% acetonitrile and applied onto a RP-HPLC preparative Nucleodur $C_{18}$(ec) 250 × 21 mm column combined with an Agilent 1100 HPLC-system. Elution was performed by application of following solvent gradient of water/0.05% formic acid (solvent A) and methanol/0.05% formic acid (solvent B) at a flow rate of 16 mL/min: linear increase from 0% to 50% B within 50 min followed by a linear increase to 95% B in 5 min, holding B for an additional 5 min. The wavelengths chosen for detection were 215 nm and 280 nm, respectively. Siderophore containing fractions were confirmed by using the CAS-assay. Positive fractions were lyophilized and subjected to further analysis. The retention time of erythrochelin was 30.7 min.

### 4.9 Methods for Stereochemical Analysis

#### 4.9.1 Stereochemical Analysis of Erythrochelin via FDAA-Derivatization

Derivatization of amino acids with Marfey's reagent (FDAA) [*N*-α-(2,4-dinitro-5-fluorophenyl)-L-alaninamide] allows the discrimination between L- and D-isomers due to altered retention times (Figure 4.5). FDAA derivatives of D-amino acids exhibit strong intramolecular bonding which reduces their polarity relative to the corresponding L-amino acids. Consequently, the D-derivatives are selectively retained on RP columns. FDAA-derivatization of erythrochelin hydrolysate was conducted to elucidate the stereochemical properties of the single amino acids.

Five hundred micrograms of erythrochelin were completely hydrolyzed by addition of 400 µL of 6 N HCl and incubation at 110°C for 24 h. The solution was lyophilized and the remaining residue dissolved in

10 µL of 1 M $NaHCO_3$. 170 µL of 1% FDAA in acetone were added and the solution was heated at 37°C for 1 h. The derivatization reaction was terminated by addition of 20 µL 1 N HCl. After lyophilization the derivatized amino acids were resolubilized by addition of 1:1 water:acetonitrile solution and 0.1% TFA to obtain a final volume of 400 µL. Products of derivatization were analyzed by RP-LCMS on a Synergi Fusion-RP 80 250 x 2.0 mm column (Phenomenex) utilizing the following solvent gradient of buffer A (10 mM ammonium formate, 1% methanol, 5% acetonitrile, pH 5.2) and buffer B (10 mM ammonium formate, 1% methanol, 60% acetonitrile, pH 5.2) at a flow rate of 0.3 mL/min: linear increase from 0% to 30% B within 30 min followed by a linear increase to 95% B in 2 min, holding B for an additional 5 min. The wavelength chosen for detection was 340 nm and the column temperature was 20°C.[206] 10 µL of sample was added to 90 µL of water prior to injection of 10 µL. To determine the stereochemistry of the present amino acids, amino acid standards (D/L-Ser and L-hOrn) were prepared to compare retention times, MS-spectra and to perform coelution experiments. The FDAA-derivatized amino acids were synthesized by incubation of 25 µL 50 mM amino acid in water, 50 µL 1% FDAA in acetone and 10 µL of 1 M $NaHCO_3$ at 37°C for 1 h. The solution was lyophilized, and the dried products resolubilized in 1:1 water:acetonitrile solution and 0.1% TFA to obtain 200 µL. L-hOrn was synthesized chemically according to an established protocol.[207] Coelution experiments were conducted by mixing 10 µL of derivatized erythrochelin hydrolysate with 1 µL of derivatized D-Ser amino acid standard and 3 µL of derivatized L-hOrn standard. RP-LCMS analysis was performed as described above.

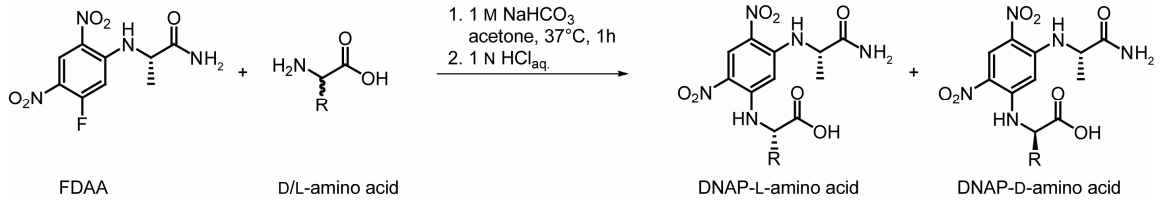



**Figure 4.5:** FDAA-derivatization of D/L-configured amino acids: The base assisted nucleophilic attack of the α-$NH_2$ group of the amino acids onto the fluorinated aryl moiety results in DNAP-modified diastereomers which can be subsequently separated on RP columns.

### 4.9.2 Determination of Erythrochelin Amino Acid Connectivity

The determination of erythrochelin amino acid connectivity together with the results obtained from amino acid stereochemistry analysis establishes the overall stereochemistry of erythrochelin. This approach took advantage of the *C*-terminal dipeptide consisting of two modified ornithines solely, which could be generated, isolated, hydrolyzed and derivatized.

3 mg of erythrochelin were partially hydrolyzed in 200 µL of 6 N HCl at 110°C for 20 min. The resulting solution was lyophilized and resolubilized in 1:1 water:acetonitrile solution and 0.1% TFA to a final volume of 200 µL and analyzed via an LTQ-FT instrument to a microbore Agilent 1100 HPLC system. Products were analyzed on a Nucleodur $C_{18}$(ec) 125 x 2 mm column, utilizing the following solvent gradient: 0-30 min, 0%-100% acetonitrile into water, both supplemented with 0.1% TFA and holding 100% acetonitrile for an additional 5 min. The column temperature was 45°C and the flow-rate was 0.3 mL/min. CID fragmentation studies within the linear ion trap were done using online LCMS. The target fragment was isolated from the mixture with an Agilent 1100 HPLC-system connected to an AnalytFC fraction collector (Agilent) on a Hypercarb 100 x 2.1 mm column (Thermo) utilizing the following solvent gradient of buffer A (10 mM ammonium formate, 1% methanol, 5% acetonitrile, pH 5.2) and buffer B (10 mM ammonium formate, 1% methanol, 60% acetonitrile, pH 5.2) at a flow rate of 0.3 mL/min: linear increase from 0% to 30% B within 30 min followed by a linear increase to 95% B in 2 min, holding B for an additional 5 min. The wavelength chosen for detection was 340 nm and the column temperature was 20°C.[206] Product containing fractions were identified by RP-LCMS. Positive fractions were lyophilized, hydrolyzed and derivatized with FDAA as described above. Analysis of the derivatized amino acids was performed via RP-LCMS.

### 4.10 Bioactivity Assays

#### 4.10.1 DNA-Bisintercalation Activity Assay

The DNA-bisintercalation activity assay is based on the capability of specific natural products to intercalate into duplex DNA. This intercalation leads to a stabilization which affects the melting temperature ($T_m$) of duplex DNA which can be measured spectrophotometrically. To investigate the intercalative properties of thiocoraline and its chemoenzymatically generated analogues, a set of oligonucleotides was synthesized (Table 4.3). The sequences of the utilized oligonucleotides were based on results obtained earlier.[134] Annealing of each 5'-oligonucleotide with its complementary 3'-oligonucleotide at a final duplex concentration of 2 µM in 10 µL phosphate buffer (10 mM sodiumphosphate, 100 mM NaCl, pH 7.0) was accomplished in a standard thermocycler (Eppendorf) by heating the solution to 95°C for 5 min and gradually cooling to 20°C at a rate of 1°C/min. Samples were incubated with authentic thiocoraline or the isolated macrocycles at a final concentration of 2 µM bisintercalator in 5 µL DMSO for 1 h at 37°C. Subsequently, phosphate buffer was added to a final volume of 200 µL and the mixture was transferred into a UV-cuvette and overlaid with silicon oil. DNA-

melting experiments were carried out on a DU-800 spectrophotometer (BeckmannCoulter). The initial temperature of 25°C was gradually increased by a ramp rate of 1°C/min to a final temperature of 95°C. Consecutively, a reverse experiment was conducted by decreasing the temperature from 95°C to 25°C utilizing the same temperature gradient. Throughout the process absorption at 260 nm was measured. The midpoint of the transition ($T_m$) was calculated and compared to a control lacking the bisintercalator. Based upon these results the stabilization of duplex DNA ($\Delta T_m$) was calculated.

**Table 4.3:** The utilized oligonucleotides for the evaluation of DNA-bisintercalation activity of thiocoraline and the chemoenzymatically generated analogues. Putative intercalation sites are accentuated in red.

| oligonucleotide | sequence (5'→3') |
|---|---|
| 5'-AT | CAATTAAATATAAC |
| 3'-AT | GTTATATTTAATTG |
| 5'-AS | AATATACGTTCGATTAA |
| 3'-AS | TTAATCGAACGTATATT |
| 5'-GC | GCGCGGCGTCCGGGCC |
| 3'-GC | GGCCCGGACGCCGCGC |
| 5'-TCGA | AATATAAATCGATAAATTAA |
| 3'-TCGA | TTAATTTATCGATTTATATT |
| 5'-GCGG | AATATAAAGCGGTAAATTAA |
| 3'-GCGG | TTAATTTATCCGCTTTATATT |

### 4.10.2 ACE Inhibition Assay

The determination of ACE inhibition by erythrochelin and *ferri*-erythrochelin was carried out according to established protocols.[208] ACE catalyzes the proteolytic cleavage of the synthetic substrate HHL which represents an analogue of the natural substrate angiotensin I (Figure 4.6). Products of proteolysis are hippuric acid (Hip) ($\lambda_{max}$ = 228 nm) and L-histidyl-L-leucine (HL). ACE inhibition assays were carried out by incubating 50 µL of HHL solution (5 mM) and 12.5 µL of inhibitor in buffer (50 mM HEPES, 300 mM NaCl, pH 8.3) at 37°C for 10 min. Subsequently, 10 µL of ACE solution (0.05 U/mL $ddH_2O$) were added and the mixture was incubated at 37°C for 2 h. The reaction was stopped by addition of 62.5 µL of 1 N HCL. The assays were analyzed by RP-LCMS on a Nucleodur $C_{18}$(ec) 125 × 2 mm column (Macherey & Nagel) utilizing the following solvent gradient of water/0.1% TFA (solvent A) and acetonitrile/0.1% TFA (solvent B) at a flow rate of 0.2 mL/min: linear increase from 0% to 95% B within 20 min holding B for an additional 5 min. The wavelength chosen for detection of released hippuric acid was 228 nm and the column temperature was 20°C.[208]

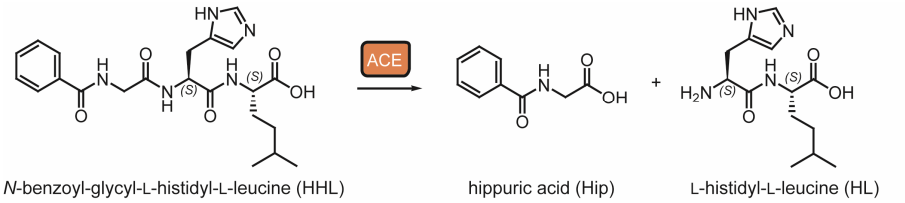



**Figure 4.6:** ACE-mediated proteolytic cleavage of the synthetic substrate HHL giving rise to the corresponding dipeptide (HL) and hippuric acid (Hip).

### 4.11 Bioinformatic Methods

Genome analysis and operon visualization was performed with Artemis, as provided by the Sanger Institute.[209] Adenylation domain specificity prediction of NRPSs was carried out with the NRPSpredictor, an online bioinformatic analysis tool.[185] The tool extracts the residues defining the active site of adenylation domains and compares the found specificity motif with known and characterized adenylation domains. In addition to analysis of the ten amino acid code defined by Stachelhaus *et al.*, residues that are positioned 8 Å around the substrate are also identified, leading to a more precise specificity prediction.[210] Genome-genome BLAST comparison was conducted with the Artemis Comparison Tool (ACT) using the tBLASTx algorithm with the E-value set to 10.0 and a BLOSUM62 scoring matrix.[211] Genome comparisons were calculated with DoubleACT using the annotated genomes of *S. erythraea* NRRL 23338 and *Shigella flexneri* 5 str. 8401 (Microbial Genome Center of ChMPH, GenBank accession number: CP000266) and a tBLASTx algorithm.[212] GeneMarkS, developed at the Georgia Institute of Technology (*Atlanta, Georgia, USA*), served as the general platform for additional genome annotation.[213]

# 5. Results

## 5.1 TE-Mediated Iterative Assembly of Chromodepsipeptides

The class of biologically active peptides is a rich resource for the discovery of novel bioactive agents. Among this class, peptides of nonribosomal origin represent structurally and functionally diverse natural products with a broad pharmacological spectrum, that are applied as antineoplastic agents (bleomycin), antibacterial compounds (daptomycin) or immunosuppressants (cyclosporine) (Table S1). This structural diversity is a result of building block variety and biosynthetic logic being either linear, nonlinear or iterative. Iterative assembly of NRPs is generally mediated by *C*-terminal thioesterase domains capable of oligomerizing linear peptidyl substrates and the modular template that is used repeatedly. In this study the cyclodimerization potential of the thioesterase domain of the thiocoraline biosynthetic machinery was investigated to obtain further insights into the iterative assembly of quinoline- and quinoxaline-type chromodepsipeptides (Figure 5.1). The biocombinatorial potential of the recombinant enzyme as well as the substrate specificity of TioS PCP-TE were determined and macrocyclization reactions were optimized to obtain maximum yields. Chemoenzymatically generated macrocycles were isolated, investigated towards DNA-bisintercalation activity and compared to native thiocoraline. In summary, TioS PCP-TE represents a robust and versatile catalyst for the generation of chromodepsipeptide analogs with potentially improved pharmacological properties.*

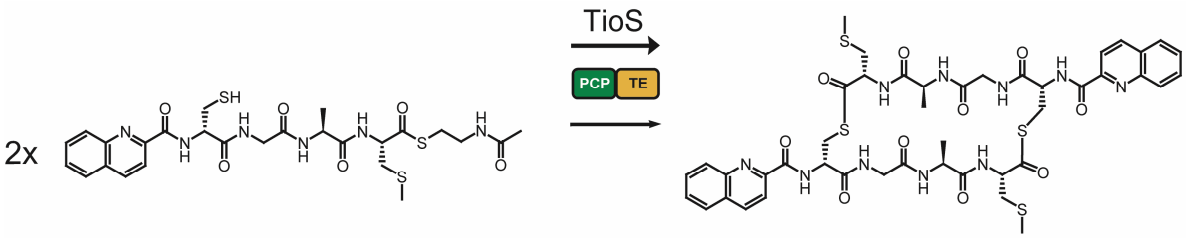



**Figure 5.1:** Chemoenzymatic approach for the generation of thiocoraline analogues based on *C*-terminal SNAC-activated tetrapeptidyl thioesters and the recombinant cyclase TioS PCP-TE. Solid-phase peptide synthesis (SPPS) affords the peptide subtrates whereas the thioesterase mediates ligation and subsequent cyclization of the substrate *in vitro*.

* The work described in chapter 5.1.1, in particular the expression and preliminary characterization of TioS PCP-TE as an active *apo*-form protein was to some extent published previously in the diploma thesis: L. Robbel, *In vitro*-Charakterisierung der Makrothiolaktonisierung von substratanalogen Peptidylthioestern durch die rekombinante Thioesterasedomäne aus der nichtribosomalen Thiocoralin Synthetase TioS, Marburg 2007. Some parts presented in the following paragraph might therefore be redundant to equivalent parts of the diploma thesis. The paragraphs were included to improve overall clarity and to present a comprehensive characterization of TioS PCP-TE.

### 5.1.1 Recombinant Expression and Isolation of TioS PCP-TE as an Active *Apo*-Protein

The gene fragment *tioS PCP-TE* was synthesized by EZBiolabs with codon usage optimization for heterologous production in *E. coli*. The fragment encoded the excised *C*-terminal PCP-TE didomain of the thiocoraline synthetase TioS, covering the *N*-terminal 40 aa upstream of the core-T motif GGSSL. *TioS PCP-TE* was cloned into the pQE60 vector and heterologously expressed in *E. coli* M15/pREP4 cells and isolated as a *C*-terminally $His_6$-tagged *apo*-form protein (Figure 5.2 A) as described in the Methods section (4.2.2). Final protein yield per liter culture after concentration and dialysis was 8.0 mg. The inclusion of the adjacent PCP-domain is assumed to ensure the correct *N*-terminal fold of the protein. Overall α-helical protein fold, as predicted for α/β-hydrolases, was confirmed via CD spectropolarimetric analysis (4.4.2) (Figure 5.2 B). The iteratively working thioesterase GrsB TE, which exhibits an analogous structural fold, served as a control (Figure 5.2 C).

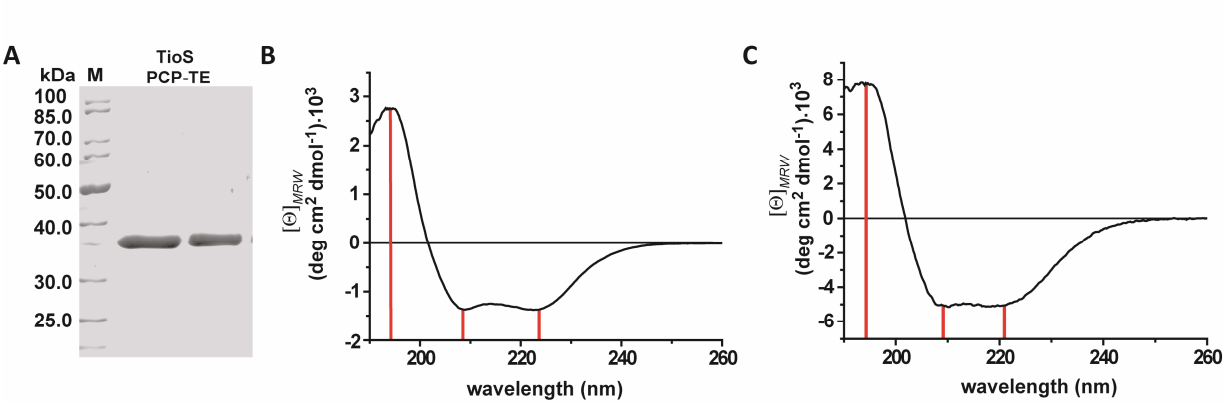



**Figure 5.2:** A) SDS-PAGE of purified recombinant TioS PCP-TE (38.5 kDa). Protein Marker (M) was Fermentas PageRuler. B/C) CD-spectra of TioS PCP-TE (B) and GrsB TE (C) as iteratively working representatives of the α/β-hydrolase family. Minima at 209 nm and 223 nm as well as the maximum at 195 nm indicate a clear overall α-helical fold of the protein being in full agreement with the proposed fold of the α/β-hydrolases.

To confirm the functionality of the catalytic triad consisting of Ser-His-Asp, which results in a deprotonated active-site serine, the activity-based fluorescence labeling assay was carried out. The recombinant proteins TioS PCP-TE, GrsB PCP-TE and GrsB TE were incubated with the activity-based probe FP-TAMRA (4.6.2). Controls were heat-denatured prior to the assay to exclude unspecific labeling of the proteins. Native proteins displayed fluorescence, due to covalent modification of the active-site serine with the fluorescent probe, whereas the denatured thioesterases were not labeled (Figure 5.3).

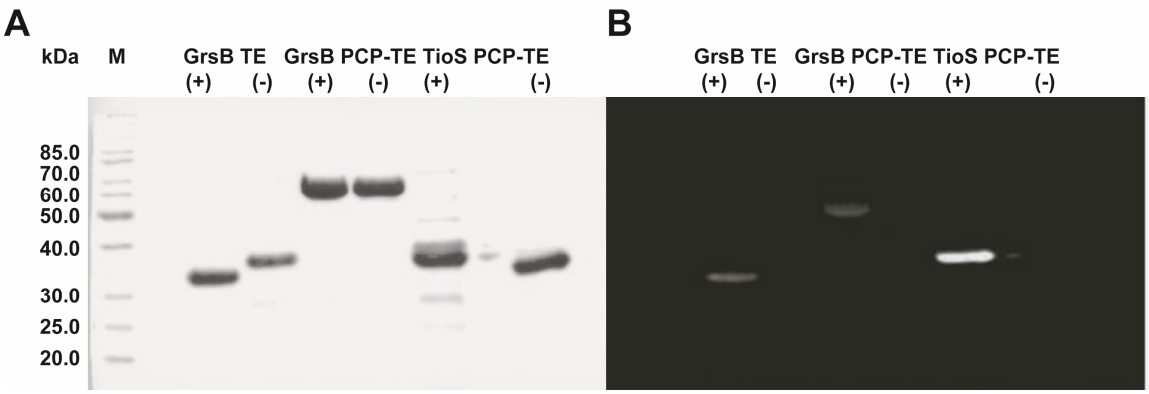



**Figure 5.3:** A) SDS-PAGE analysis of GrsB TE, GrsB PCP-TE and TioS PCP-TE after FP-TAMRA labeling studies. B) SDS-PAGE analysis of labeled proteins prior to Coomassie staining on an UV-screen ($\lambda$ = 312 nm). Labeled proteins that contain an active-site serine display fluorescence due to covalent modification with the fluorescent probe. Native thioesterases are indicated with (+) and denatured thioesterases with (-). The protein marker was Fermentas PageRuler.

### 5.1.2 Substrate Specificity of TioS PCP-TE

To evaluate the biocombinatorial potential and to investigate the combined ligation and macrocyclization mechanism of the excised TE-domain TioS PCP-TE, a set of tetrapeptidyl-thioesters was synthesized and incubated with the recombinantly generated protein (10 µM) (4.5.1). The sequences of the tetrapeptidyl-substrates were initially based on the primary amino acid sequence of the linear thiocoraline tetrapeptidyl-precursor (Figure 5.4). To overcome the lack of synthetically demanding building blocks for solid phase peptide synthesis (SPPS) and to allow the generation of novel thiocoraline analogs, naturally occurring modified amino acids were substituted with commercially available ones. The utilized substrates lacked *N*-methylation of the *C*-terminal cystein-residues and the 3-hydroxyfunctionality of the chromophore moiety 3-hydroxy-quinaldic acid (3HQA), whereas stereochemical information was conserved throughout the oligopeptide chain. For stability reasons the tetrapeptidyl substrates were *C*-terminally activated as SNACs circumventing thiophenol-activation (4.5.2). Furthermore, synthetically demanding synthesis of the octapeptidyl-precursors was circumvented by the ligation capability of TioS PCP-TE resulting in the utilization of tetrapeptidyl-precursors.

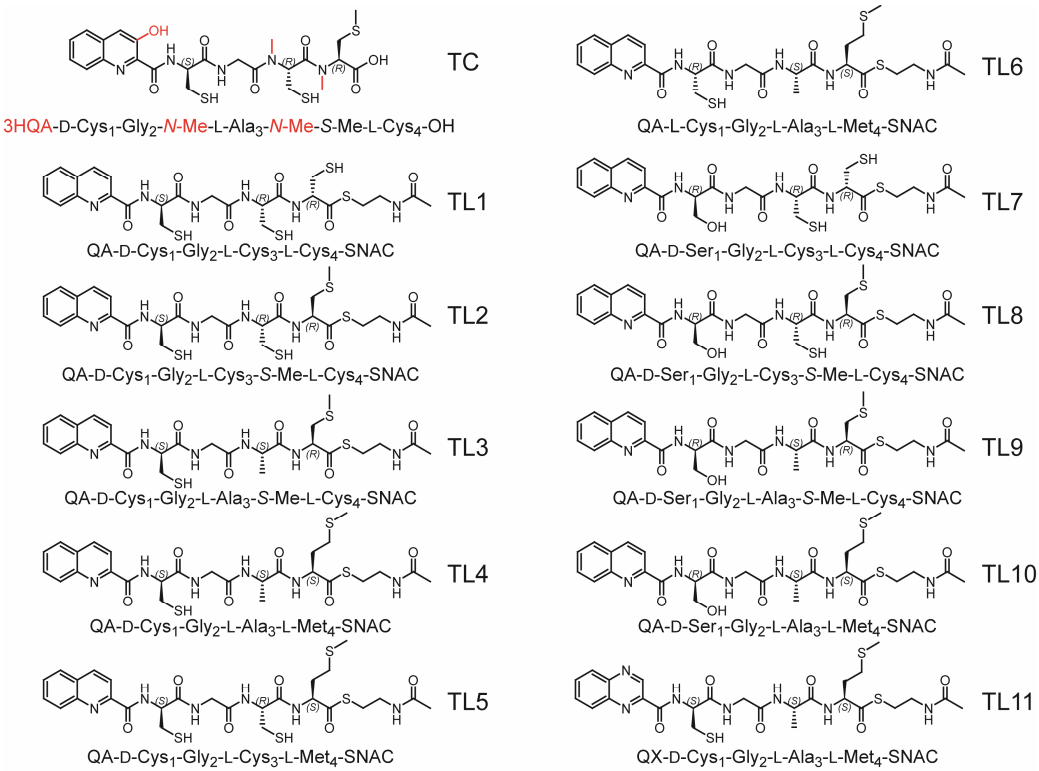



**Figure 5.4**: The chemical structures and sequences of the synthesized tetrapeptidyl SNAC-thioesters. The native thiocoraline tetrapeptide (TC) is given. All substrates employed throughout the studies lacked the hydroxylation of the chromophore moiety and *N*-methylations of the peptidic backbone (accentuated in red).

All assays were analyzed via RP-LCMS methods. Incubation of recombinantly produced TioS PCP-TE with TL1, resembling the most native substrate based on NRPS A-domain specificity prediction, revealed solely hydrolytically cleaved linear tetrapeptide (4.6.3). After 1 h total substrate hydrolysis was detected. This result indicated that the steric demand or the polarity of the *C*-terminal amino acid is essential for recognition of the substrate and subsequent ligation and macrocyclization. In addition, it was speculated that *S*-Me-L-$Cys_4$ is incorporated into the oligopeptide chain instead of L-$Cys_4$. This model of biosynthesis would require *S*-methylation prior to cyclization *in vivo*. Substitution of L-$Cys_4$ with the sterically more demanding *S*-Me-L-$Cys_4$ (TL2) also led to the exclusive formation of hydrolytically cleaved linear tetrapeptide. Based on these result,s L-$Cys_3$ was replaced with L-$Ala_3$ (TL3) in order to maintain stereochemical information and to minimize electrostatic repulsion effects between sulfhydryl groups in close proximity. HPLC-MS-analysis of the assay revealed the formation of macrothiolactone Cy3 ($t_R$ = 27.3) with a hydrolysis (Hy3, $t_R$ = 12.1) to cyclization-ratio of 12:1 and total substrate conversion after 2 h at 25°C (Figure 5.5).

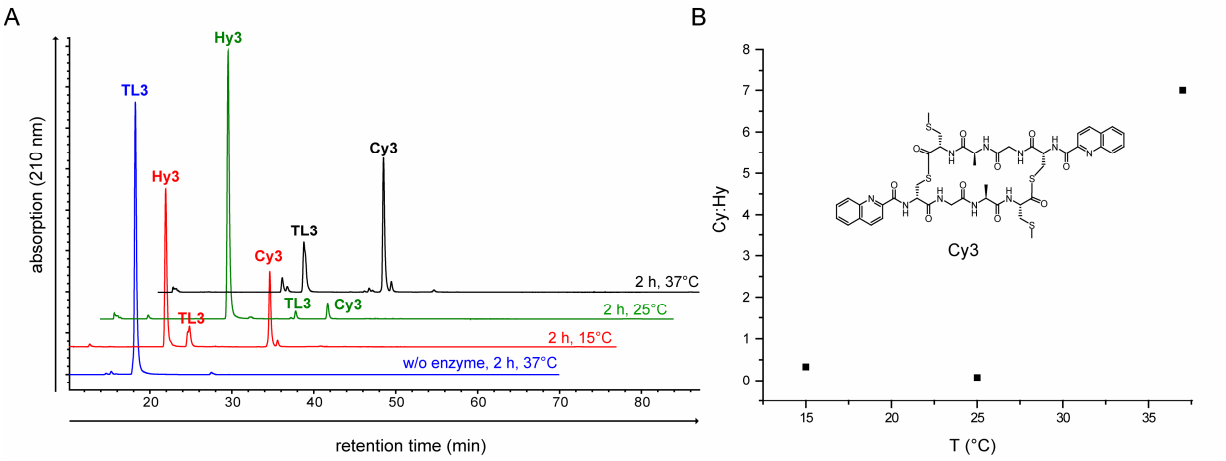



**Figure 5.5**: Cyclization of substrate TL3 mediated by TioS PCP-TE. A) The HPLC traces correspond to the incubation of TL3 (300 µM) with TioS PCP-TE at specific temperatures for 2 h. The blue HPLC trace corresponds to the control lacking the enzyme. Enzymatic conversion of TL3 leads to the hydrolytically cleaved tetrapeptide Hy3 at $t_R$ = 12.1 and the macrothiolactone Cy3 at $t_R$ = 27.3. B) Cyclization to hydrolysis-ratios depending on the reaction temperature and the chemical structure of Cy3. Ratios are based on HPLC-MS-analyses of the assays after 2 h incubation at the specified temperature.

Encouraged by the results obtained and to investigate the mechanism of macrocyclization, the steric demand of the *C*-terminal amino acid was further increased by the incorporation of L-$Met_4$ (TL4) showing an improved hydrolysis to cyclization-ratio of 1:2 (Hy4, $t_R$ = 14.2; Cy4, $t_R$ = 24.7). Additionally, the formation and accumulation of the linear octapeptidyl-SNAC (Lig4, $t_R$ = 24.1) was observed representing the main product. In total, substrate TL4 was converted at a ratio of 1:4:2 (Hy4/Lig4/Cy4) (Figure 5.6). The steric demand of the *C*-terminal L-$Met_4$ led to the covalent trapping of the ligation product and abolished complete macrocyclization. To corroborate the result that L-$Cys_3$ strongly affects macrothiolactonization, TL5 was synthesized showing a mixed substitution pattern of L-$Cys_3$ and L-$Met_4$. Analogously to the results obtained with TL2, TioS PCP-TE is not capable of catalyzing ligation or macrothiolactonization. Total substrate turnover is accomplished after 2 h resulting in complete hydrolytic cleavage of the thioester.

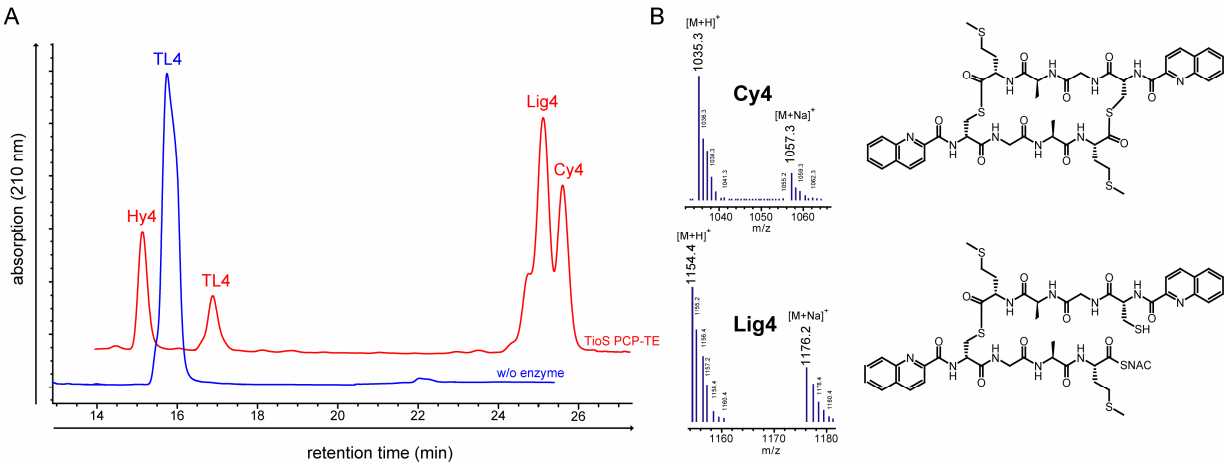



**Figure 5.6:** Cyclization and ligation of substrate TL4 mediated by TioS PCP-TE. A) The red HPLC trace corresponds to the incubation of TL4 (300 µM) with TioS PCP-TE at 25°C for 2 h. The blue HPLC trace corresponds to the control lacking the enzyme. Enzymatic conversion of TL4 leads to the hydrolytically cleaved tetrapeptide Hy4 at $t_R$ = 14.2 and the macrothiolactone Cy4 at $t_R$ = 24.7. The sterical demand of the *C*-terminal L-Met residue gives rise to the formation of the ligation product (Lig4, $t_R$ = 24.1) B) MS-spectra and chemical structures of the products: Cy4 ($[M+H]^+$ = 1035.3, $[M+Na]^+$ = 1057.3), Lig4 ($[M+H]^+$ = 1154.4, $[M+Na]^+$ = 1176.2).

All chromodepsipeptides share a D-configured amino acid in position 1, responsible for the nucleophilic attack of the side chain onto the acyl-O-TE oxoester intermediate. To prove the significance of this stereoinformation substrate TL6 was synthesized harboring L-$Cys_1$ instead of D-$Cys_1$. Using the linear tetrapeptidyl substrate solely hydrolysis was detected confirming the necessity of the *N*-terminal stereogenic center.

### 5.1.3 Biocombinatorial Evaluation of TioS PCP-TE

In order to generate novel chromodepsipeptides with improved physico-chemical stability based on the structure of thiocoraline, an alternative set of subtrates was synthesized carrying D-$Ser_1$ instead of D-$Cys_1$ as the cyclization-mediating nucleophile. Employment of TL7, the serine-substituted analogue of TL1, showed macrocylization at a hydrolysis to cyclization-ratio of 5:1. Products were assigned to the hydrolysis product (Hy7), the macrolactone (Cy7) and a macrolactone with intramolecular disulfide connectivity (Cy7SS). Consecutively, TL8 was incubated with TioS PCP-TE. After 60 min complete substrate conversion was detected with a hydrolysis (Hy8, $t_R$ = 11.2) to cyclization (Cy8, $t_R$ = 30.2)-ratio of 2:1 (Figure 5.7). Autoxidation of Cy8 gives rise to the disulfide-bridged macrolactone Cy8SS ($t_R$ = 32.7). Additionally, side-product formation could be assigned to a four residue macrolactone Cy8/4 resulting

from an intramolecular attack of the side chain nucleophile of D-Ser$_1$ onto the acyl-O-TE oxoester (Cy8/4 $t_R$ = 20.5) (Figure S1).

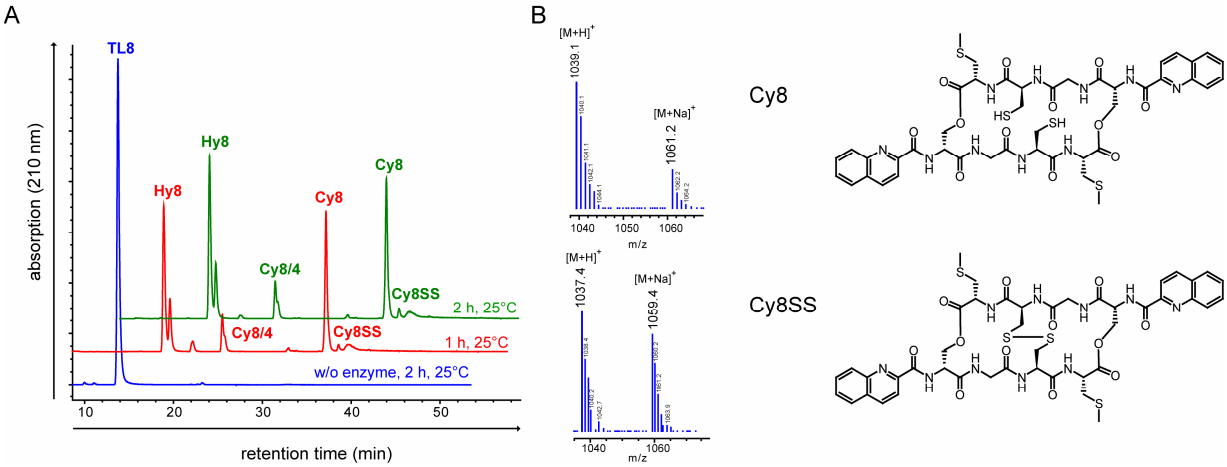



**Figure 5.7:** Cyclization and ligation of substrate TL8 mediated by TioS PCP-TE. A) The red HPLC trace corresponds to the incubation of TL8 (300 µM) with TioS PCP-TE at 25°C for 1 h and the green trace for the same reaction for 2 h. The blue HPLC trace corresponds to the control lacking the enzyme. Enzymatic conversion of TL8 leads to the hydrolytically cleaved tetrapeptide Hy8 at $t_R$ = 11.2 and the macrolactone Cy8 at $t_R$ = 30.2. Oxidative formation of an intramolecular disulfide bridge gives rise to macrolactone Cy8SS. In addition, side-product formation of Cy8/4 at $t_R$ = 30.2 was observed. B) MS-spectra and the chemical structures of the products: Cy8 ($[M+H]^+$ = 1039.1, $[M+Na]^+$ = 1061.2), Cy8SS ($[M+H]^+$ = 1037.4, $[M+Na]^+$ = 1059.4).

MS-fragmentation studies strongly support the identity of the four residue macrolactone and exclude the formation of an alternative two residue macrothiolactone, due to the detection of intense fragments containing dehydro-alanine, which are characteristic for gas-phase fragmentations of lactones (Figure S1).[214] Macrocyclization of the linear tetrapeptidyl-SNAC was exclusively limited to substrate TL8. Substitution of L-Cys$_3$ with L-Ala$_3$ and subsequent incubation of substrate TL9 with TioS PCP-TE led to a hydrolysis (Hy9 $t_R$ = 7.5) to cyclization (Cy9 $t_R$ = 27.5)-ratio of 8:1 at 25°C (Figure 5.8). Based on the results obtained with TL4, the analogous substrate TL10 was synthesized. HPLC-MS analysis revealed the formation of the macrocycle Cy10 at a hydrolysis-to-cyclization-ratio of 8:1. Additionally, the formation of the linear octapeptidyl-SNAC Lig10 was detected, reflecting the steric demand of L-Met$_4$. To investigate the influence of the chromophore moiety on cyclization-efficiency and to establish TioS PCP-TE as a general catalyst for the ligation and cyclization of the quinoline- and quinoxaline-type class of chromodepsipeptides, TL11 was synthesized. The primary sequence was based on TL3 with the exception of the chromophore moiety quinaldic acid (QA), which was substituted with quinoxaline-2-

carboxylic acid (QX), the chromophore found in echinomycin and triostin A. The cyclization reaction profile revealed a hydrolysis-to-cyclization-ratio of 8:1 in analogy to substrate conversion of TL3. Substrate conversion was completed after 1 h of incubation at 25°C. This result in general indicates the relaxed substrate specificity of the cyclase towards the *N*-terminal chromophore and implies that TioS PCP-TE can serve as a prototypical TE for the assembly of quinoline or quinoxaline carrying compounds.

### 5.1.4 Temperature Dependence of Cyclodimerization Reactions

Enzymes influence the kinetic properties of the catalyzed reaction but thermodynamic parameters remain unaffected. The formation of the macrocyclic lactone from two thioester-activated building blocks represents an energetically favored reaction, from a thermodynamic point of view, resulting in a +M-stabilized structure. In contrast, the macrothiolactone represents a thermodynamically instable product due to the low bond dissociation energy, which is the result of the missing resonance stabilization of the S-C-bond. To analyze if the formation of macrolactones and the energetically unfavored macrothiolactones can be influenced by altering the thermodynamics of the enzymatic reaction, varying temperatures were employed when conducting the assays. In order to improve the cyclization yields and to decrease the hydrolytic release of the linear peptidyl-precursor, the substrates TL3 and TL9, both differing only in the nature of the cyclization-mediating nucleophile (D-$Cys_1$ or D-$Ser_1$) were analyzed towards varying assay temperatures. The temperatures chosen were 15°C, 25°C and 37°C respectively. TL3 showed an improved hydrolysis to cyclization-ratio of 3:1 at 15°C compared to a ratio of 12:1 (Hy3/Cy3) at 25°C. The best macrothiolactone (Cy3, $t_R$ = 27.3) yields were obtained at 37°C with an altered reaction profile revealing a low flux towards hydrolysis (Hy3, $t_R$ = 12.1) and a shifted hydrolysis to cyclization-ratio of 1:7 (Figure 5.5). Kinetic parameters were determined for total substrate conversion at 37°C revealing a $k_{cat}$ of 5.26 ± 0.64 $min^{-1}$. In contrast, TL9 was cyclized more efficiently at low temperatures. An improved hydrolysis to cyclization-ratio of 4:1 was observed at 15°C compared to a ratio of 8:1 at 25°C (Hy9, $t_R$ = 7.5; Cy9, $t_R$ = 27.3) (Figure 5.8). Interestingly, cyclization was completely abolished at 37°C. Solely the formation of the linear tetrapeptide (Hy9) was detected. Additionally, substrate TL8 was examined towards temperature dependence of macrolactonization. Best macrocyclization yields were obtained at 15°C with a hydrolysis to cyclization-ratio of 1:4 compared to a ratio of 2:1 at 25°C (Hy8, $t_R$ = 11.2; Cy8, $t_R$ = 30.2) (Figure 5.7). At 37°C cyclization yields were reduced, consistent with the results obtained with TL9, to a ratio of 9:1 towards hydrolysis. Kinetic parameters were determined for TL8 at 15°C resulting in a $k_{cat}$ of 8.92 ± 1.2 $min^{-1}$.

The data obtained clearly demonstrates the applicability of this approach which allows the improvement of the investigated cyclodimerizations for maximum macrocycle yield by altering reaction temperatures.

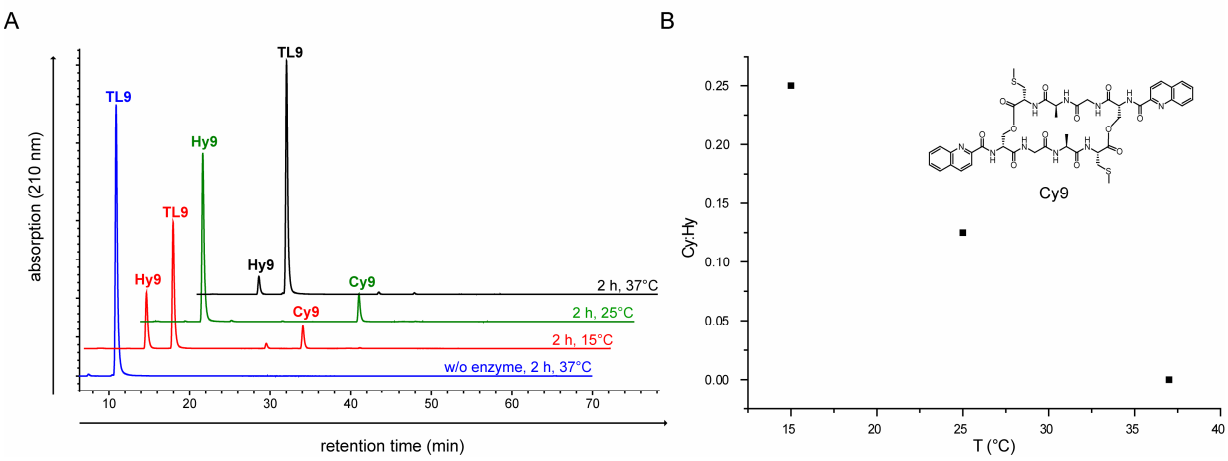



**Figure 5.8**: Cyclization of substrate TL9 mediated by TioS PCP-TE. A) The HPLC traces correspond to the incubation of TL9 (300 µM) with TioS PCP-TE at specific temperatures for 2 h. The blue HPLC trace corresponds to the control lacking the enzyme. Enzymatic conversion of TL9 leads to the hydrolytically cleaved tetrapeptide Hy9 at $t_R$ = 7.5 and the macrolactone Cy9 at $t_R$ = 27.3. B) Cyclization to hydrolysis-ratios depending on the reaction temperature and the chemical structure of Cy9. Ratios are based on HPLC-MS-analyses of the assays after 2 h incubation at the specified temperature

### 5.1.5 DNA-Bisintercalation Activities

Thiocoraline and analogous natural products attain their bioactivity as antiproliferative agents from DNA-bisintercalation properties. In order to evaluate the DNA-bisintercalative properties of the chemoenzymatically generated thiocoraline analogues and to elucidate structural features contributing to DNA-insertion, a DNA-bisintercalation assay was established (4.10.1). Initially, a set of five oligonucleotides (Table 4.3) was analyzed for the native $T_m$ based on UV-absorption (Figure 5.9).[134]

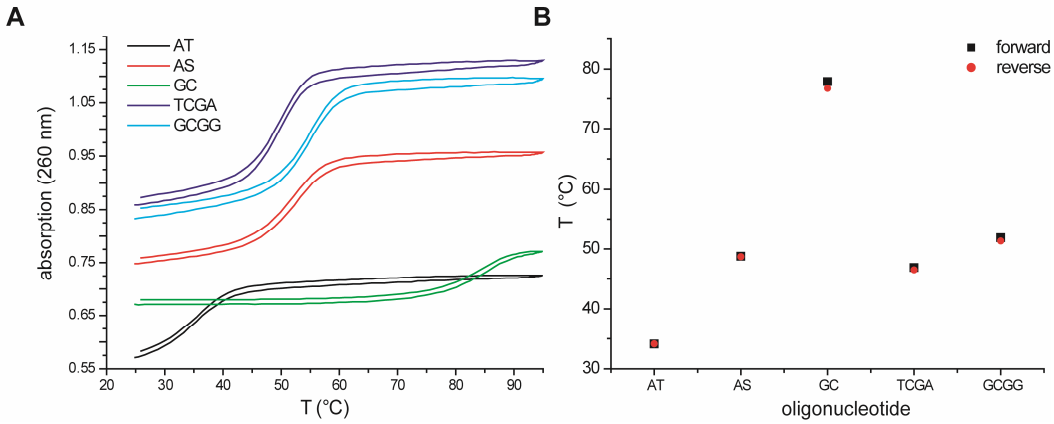



**Figure 5.9:** A) UV-absorption profiles of the investigated oligonucleotides at 260 nm. The midpoint of transition ($T_m$) was calculated for the forward and the reverse experiment, which was conducted by decreasing the temperature from 95°C to 25°C utilizing the same temperature gradient. B) The calculated reverse and forward $T_m$-values for the corresponding oligonucleotides.

To compare the bisintercalation properties of the novel analogues with known intercalation agents, native thiocoraline was isolated and subjected to the DNA-melting assay (4.8.2). Incubation of the oligonucleotides with thiocoraline and subsequent DNA-melting experiments resulted in melting curves showing a hysteresis shape characteristic for DNA-bisintercalators. All nucleotides experienced thiocoraline-induced duplex-stabilization at a concentration of 2 µM, proving the applicability of the established system for the investigation of natural product DNA-bisintercalation activity (Figure 5.10). The oligonucleotide AS displayed the highest degree of stabilization ($\Delta T_m$ = 15.9°C) and was chosen for subsequent experiments employing the chemoenzymatically generated macrocycles. To elucidate structural features contributing to DNA-insertion, four tetrapeptidyl thioesters were synthesized. In accordance with the previously discussed results the macrocyclization assays were carried out under optimal conditions. Isolation of the corresponding macrocycles (Cy3/Cy8/Cy8SS/Cy9/Cy11) was achieved by RP-HPLC separation (4.8.1). Incubation of AS with the isolated macrolactones and macrothiolactones

led only to a marginal stabilization of 0.1-0.2°C, suggesting that substitutions within the peptide backbone strongly influence macrocycle-DNA interactions (data not shown).

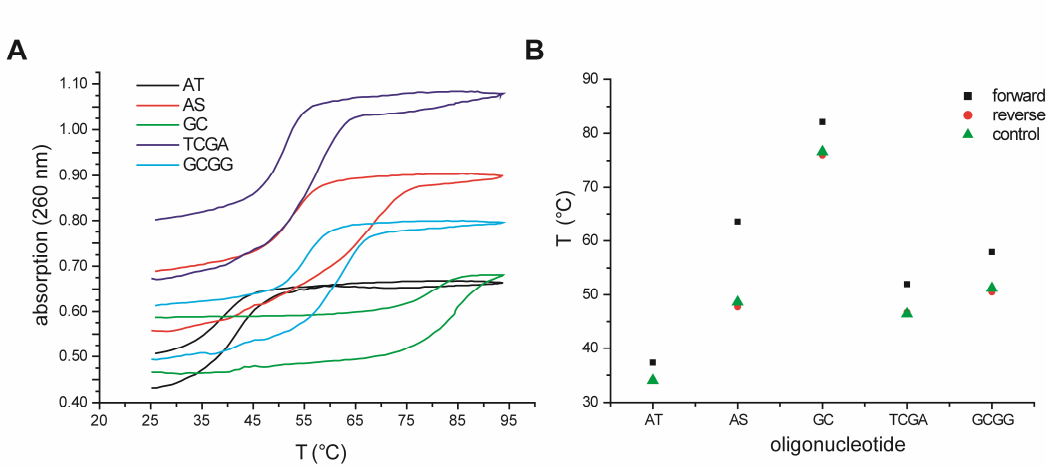



**Figure 5.10:** A) UV-absorption profiles of the investigated oligonucleotides at 260 nm after incubation with 2 µM thiocoraline. Duplex stabilization by thiocoraline resulted in typical hysteresis shaped melting curves, characteristic for DNA-bisintercalators. B) The calculated reverse and forward $T_m$-values for the stabilized oligonucleotides compared to the control lacking thiocoraline show the degree of stabilization.

## 5.2 Erythrochelin – a Hydroxamate-Type Siderophore Predicted from the Genome of *S. erythraea*

Advances in sequencing technologies gave rise to a plethora of sequenced and annotated microbial genomes. Bioinformatic analysis of these microbial DNA sequences revealed a multitude of cryptic gene clusters, demonstrating a tremendous potential for the discovery of natural products. The isolation of the corresponding products of these cryptic clusters is challenging due to either a low rate of production or unknown conditions for secondary metabolite biosynthesis. In addition, bioactivity-guided natural product isolation is often impeded by unpredictable biological activities of the target compounds and the lack of appropriate screening-methods. To circumvent the problem of a low rate of biosynthesis and unknown biological activity, a genome mining approach relying on bioinformatic genome analysis and transcriptome data combined with radiolabeled precursor feeding studies for NRPS-derived natural products was carried out. In this methodology transcriptome-analysis provides information on the growth conditions leading to gene cluster expression, whereas A-domain specificity prediction defines the radiolabeled precursor. This approach led to the discovery of erythrochelin, a hydroxamate-type siderophore, isolated from cultures of the actinomycete *S. erythraea*. Production conditions were optimized to obtain sufficient amounts for the structural and functional characterization of the compound, which relied on NMR- and $MS^n$-analysis. The overall stereochemistry was determined via derivatization and LCMS-analysis of the single amino acids and the functional properties of erythrochelin acting as an iron-chelating compound and potential inhibitor of ACE were investigated. In summary, the radio-LCMS-guided genome mining proved to be advantageous for the initial detection of NRPS-derived natural products due to the sensitivity of radioactivity detection and the sophisticated analytical separation.

### 5.2.1 Bioinformatic Identification of Two Siderophore Biosynthetic Gene Glusters in *S. erythraea*

Analysis of the sequenced and annotated genome of *S. erythraea* led to the discovery of two NRPS-gene clusters linked to siderophore biosynthesis and transport.[215] While one of the two was predicted to encode for a mixed hydroxamate/catecholate-type siderophore (*nrps3*), the second operon was envisaged to encode a tetramodular NRPS putatively capable of assembling a hydroxamate-type siderophore (*etc*). Both biosynthetic gene clusters are located in the noncore region of the *S. erythraea* genome, which is housing the majority of secondary metabolite operons.[215]

The *etc* operon contains 11 CDSs which are clustered in a region covering 28.8 kb with an average GC-content of 71.2% (Figure 5.11). The NRP synthetase encoded by *etcD* (*sace_3035*) comprises four modules, each containing the essential condensation (C), adenylation (A) and peptidyl carrier protein (PCP) domains. In addition, module 1 and module 2 contain an epimerisation (E) domain each, responsible for stereoconversion of the accepted L-amino acids to the D-isomer, indicating the presence of two D-configured residues in the assembled product.

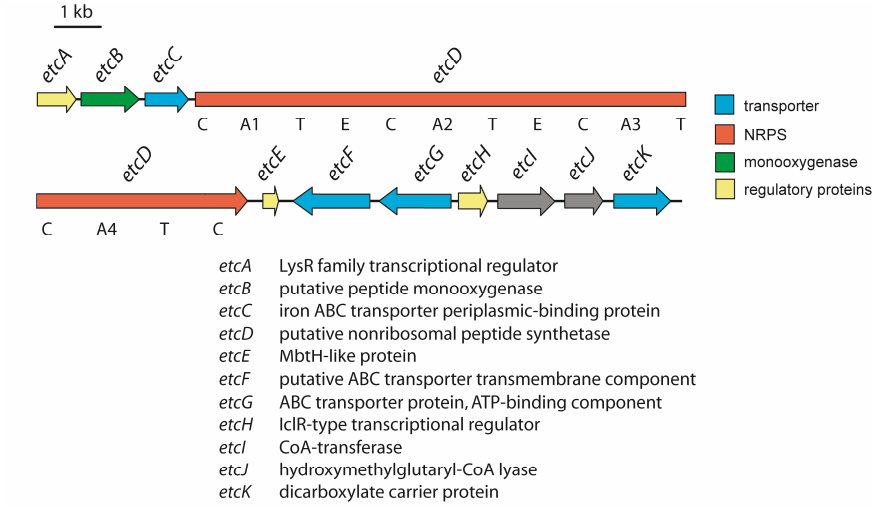



**Figure 5.11:** Schematic overview of the *etc* gene cluster. Putative functions of the proteins encoded within the operon are based on BLAST-analysis and are given in the figure. Apart from the core components for siderophore biosynthesis, genes encoding for exporters and importers of the siderophore as well as typical transcriptional regulators for secondary metabolism are found, determining the boundaries of the cluster.

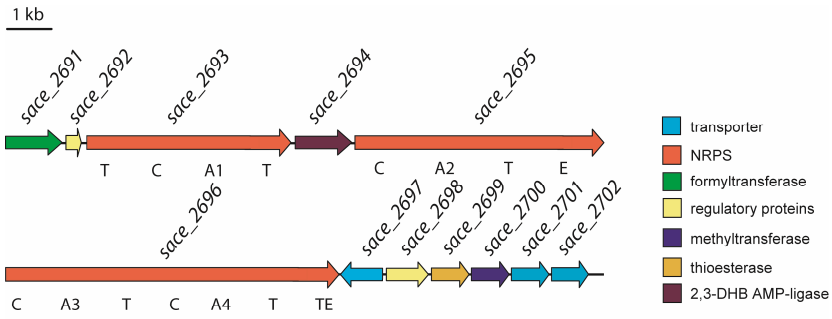
The *N*-terminal region of module 1 shares a high degree of homology to condensation domains suggesting the function of an initiation module mediating the condensation of an external building block with the PCP-tethered substrate. Module 4 contains a *C*-terminal C-domain instead of a thioesterase (TE)-domain commonly responsible for product release through hydrolytic cleavage or macrocyclization.[216] Upstream of *etcD*, a gene with high sequence homology to characterized L-ornithine hydroxylases is located. EtcB shares 64% sequential identity and 78% sequential similarity with CchB, the hydroxylase required for coelichelin biosynthesis.[98] Based on the proposed function of EtcB, the incorporation of δ-*N*-hydroxy-L-ornithine residues into the readily assembled oligopeptide was predicted. Furthermore, genes present in the cluster code for proteins typically associated with secondary metabolite biosynthesis and siderophore transport: a transcriptional regulator (*etcA*), an MbtH-like protein (*etcE*) and proteins for siderophore export and uptake (*etcCFGK*). EtcC is proposed to be a membrane-anchored ferric-siderophore binding protein involved in siderophore uptake similar to the

periplasmic binding protein CchF, putatively mediating coelichelin uptake.[38] The upstream edge of the *etc* cluster is defined by *etcA*, which encodes a regulatory protein of the LysR-family, whereas the downstream edge is defined by *etcK*, encoding a dicarboxylate carrier protein. In addition, a CoA-transferase (EtcI) and a hydroxymethylglutaryl-CoA ligase (EtcJ) are encoded within the cluster. A comprehensive bioinformatic analysis of the proteins encoded within the *etc* gene cluster is given in Table S2. The amino acid specificity of the synthetase EtcD was predicted by using a methodology comparing active-site residues of known NRPS adenylation domains with the adenylation domains found in EtcD (Table 5.1).[185,210,217] The first adenylation domain ($A_1$) is predicted to activate L-arginine but shows only 70% identity of the residues determining the specificity to MycC, suggesting the activation of a structurally analogous building block. MycC itself represents a NRPS-termination module involved in the assembly of microcystin by *Microcystis aeruginosa* PCC7806, predicted to activate L-arginine.[218] $A_2$ and $A_3$ are predicted to activate L-serine and δ-*N*-hydroxy-L-ornithine (L-hOrn), respectively, as found in the assembly of enterobactin (EntF) and coelichelin (CchH).[38,219] The *C*-terminal adenylation domain $A_4$ again is predicted to activate L-arginine, displaying 60% identity to the characterized A-domain of MycC. Interestingly, $A_1$ and $A_4$ feature a highly identical (90%) specificity determining residue pattern, leading to the assumption that both activate the same substrate. Based on the bioinformatic analysis of the *etc* gene cluster it was predicted that the assembled tetrapeptide consists of L-hOrn, L-Ser and two building blocks analogous to L-Arg.

**Table 5.1:** Comparison of the extracted active-site residues determining the adenylation domain specificity of EtcD with known adenylation domains. Variations in the residue pattern are highlighted in red. The substrate prediction for each A-domain as well as the product of the NRPS is given. MycC, microcystin synthetase; EntF, enterobactin synthetase; CchH, coelichelin synthetase.

| **predicted adenylation domain specificity** | | | |
|---|---|---|---|
| A domain | active site residues | substrate | product |
| $A_1$ | D V W A L G A V N K | | |
| MycC | D V W T I G A V D K | L-Arg | microcystin |
| $A_2$ | D V W H F S L V D K | | |
| EntF | D V W H F S L V D K | L-Ser | enterobactin |
| $A_3$ | D M E N L G L I N K | | |
| CchH-$A_3$ | D M E N L G L I N K | L-hOrn | coelichelin |
| $A_4$ | D V F A L G A V N K | | |
| MycC | D V W T I G A V D K | L-Arg | microcystin |

The *nrps3* biosynthetic operon contains 12 CDSs which are clustered in a region covering 25.5 kb with an average GC-content of 71.4%. It encompasses three nonribosomal peptide synthetases (*sace_2693, sace_2695, sace_2696*) which together consist of four individual modules, each containing all essential domains for peptide elongation (Figure 5.12). The *N*-terminal region of the first synthetase Sace_2693 shares a high degree of homology to PCP-domains and is putatively involved in the initiation of siderophore biosynthesis. In this model, the 2,3-DHB AMP-ligase Sace_2694, which is homologous to DhbE, catalyzes the activation of 2,3-DHB as an AMP-derivative.[220] This intermediate is subsequently transferred onto the 4'-ppan group of Sace_2693 and initiates tetrapeptide assembly. An E-domain located at the *C*-terminus of Sace_2695 is postulated to mediate the conversion of the initially-activated L-amino acid to the D-isomer. The dimodular NRPS Sace_2696 is proposed to be involved in the assembly and subsequent release of the tetrapeptide by the TE-domain located at the *C*-terminus of the synthetase.




| | |
|---|---|
| *sace_2691* | methionyl-tRNA formyltransferase |
| *sace_2692* | MbtH-like protein |
| *sace_2693* | nonribosomal peptide synthetase |
| *sace_2694* | 2,3-DHB AMP-ligase |
| *sace_2695* | putative nonribosomal peptide synthetase |
| *sace_2696* | putative nonribosomal peptide synthetase |
| *sace_2697* | ABC-type $Fe^{3+}$-siderophore transporter |
| *sace_2698* | putative regulatory protein |
| *sace_2699* | thioesterase |
| *sace_2700* | O-methyltransferase |
| *sace_2701* | ABC-transporter |
| *sace_2702* | polyketide exporter |

**Figure 5.12:** Schematic overview of the *nrps3* gene cluster. Putative functions of the proteins encoded within the operon are based on BLAST-analysis and are given in the figure. Apart from the core components for siderophore biosynthesis, genes encoding for exporters and importers of the siderophore as well as typical transcriptional regulators for secondary metabolism are found. The boundaries are determined by a gene encoding a putative formyltransferase and a gene encoding an ABC-type transporter.

The upstream edge of the cluster is defined by sace_2691, encoding a methionyl-tRNA formyltransferase. A common modification of hydroxamate-type siderophores is the formylation of hydroxylated lysine or ornithine residues that is belived to be carried out by formyltransferases utilizing

the cosubstrate $N^{10}$-f$H_4$F.[221-222] The presence of a formyltransferase in the biosynthetic cluster indicates the incorporation of a formylated building block into the tetrapeptide sequence. Additional structural diversity is putatively introduced by *O*-methylation of specific residues, mediated by the *O*-methyltransferase Sace_2700.[223] The downstream edge of the gene cluster is defined by the ABC-type exporter encoding genes *sace_2701* and *sace_2702*. Further genes located within the cluster code for a MbtH-like protein (*sace_2692*), a periplasmic siderophore binding protein (*sace_2797*) and a lone-standing type II thioesterase (*sace_2699*).

Analysis of the specificity-conferring residues was conducted utilizing the methods mentioned above (Table 5.2).[185,210,217] The first adenylation domain is predicted to activate L-Orn as it shows 70% identity to the specificity motif of FxbC-$A_1$, involved in exochelin assembly.[224] $A_2$ is predicted to activate L-Asn and the cognate subtrate for $A_3$ is predicted to be L-Thr as found in the calcium-dependent antibiotic (CDA).[187] The fourth adenylation domain $A_4$ is proposed to activate L-hfOrn and contains a specificity determining residue pattern which is identical to CchH-$A_1$, responsible for coelichelin biosynthesis.[38] The incorporation of L-hfOrn is in full agreement with the formyltransferase Sace_2691, encoded in the *nrps3* gene cluster. Based on the analysis of A-domain specificities and the putative functions of the enzymes encoded within the cluster, it is predicted that the hydroxamate/catecholate-type siderophore consists of L-Orn, D-Asn, L-Thr and L-hfOrn. The *N*-terminus of the tetrapeptide is capped with the catecholate moiety 2,3-DHB, which also participates in iron-chelation.[146]

**Table 5.2:** Comparison of the extracted active-site residues determining the adenylation domain specificity of Sace_2693, Sace_2695 and Sace_2696 with known adenylation domains. Variations in the residue pattern are highlighted in red. The substrate prediction for each A-domain as well as the product of the NRPS is given. FxbC, exochelin synthetase; CDA PSII-I, CDA synthetase; CchH, coelichelin synthetase.

| | **predicted adenylation domain specificity** | | |
|---|---|---|---|
| A domain | active site residues | substrate | product |
| $A_1$ | D L F N L G L I H K | | |
| FxbC-$A_1$ | D M E N L G L I N K | L-Orn | exochelin |
| $A_2$ | D F T K V A E V G K | | |
| CDA PSII-$A_3$ | D L T K V G E V G K | L-Asn | CDA |
| $A_3$ | D F W N V G M V H K | | |
| CDA PSI-$A_2$ | D F W N V G M V H K | L-Thr | CDA |
| $A_4$ | D I N Y W G G I G K | | |
| CchH-$A_1$ | D I N Y W G G I G K | L-hfOrn | coelichelin |

### 5.2.2 Identification and Isolation of Erythrochelin via Radio-LCMS-Guided Genome Mining

Transcriptome data of *S. erythraea* using GeneChip DNA microarrays, collected by De Bellis and coworkers, indicates up-regulation of gene expression associated with the *etc* and the *nrps3* gene clusters under specific conditions.[225] It was shown that gene expression of *nrps3* was upregulated in growth phase B (30-50 h) and *etc* gene cluster expression was upregulated in the transition phase B/C (45-65 h). To identify and isolate the putative siderophores, cultures of *S. erythraea* were grown under the same conditions as employed for collection of transcriptome data and growth phase was monitored by LCMS-analysis using erythromycin production as the indicator of the growth phase. An extraction of the SCM-medium supernatant after 4 d of growth, subsequent preparative HPLC fractionation and CAS-liquid assay analysis of the fractions revealed a CAS-reactive compound (Chapter 4.8.3) (Figure 5.13).[205] The coelution of a multitude of compounds in the CAS-assay positive fraction impeded the direct MS-based detection and isolation of the siderophore due to media complexity.

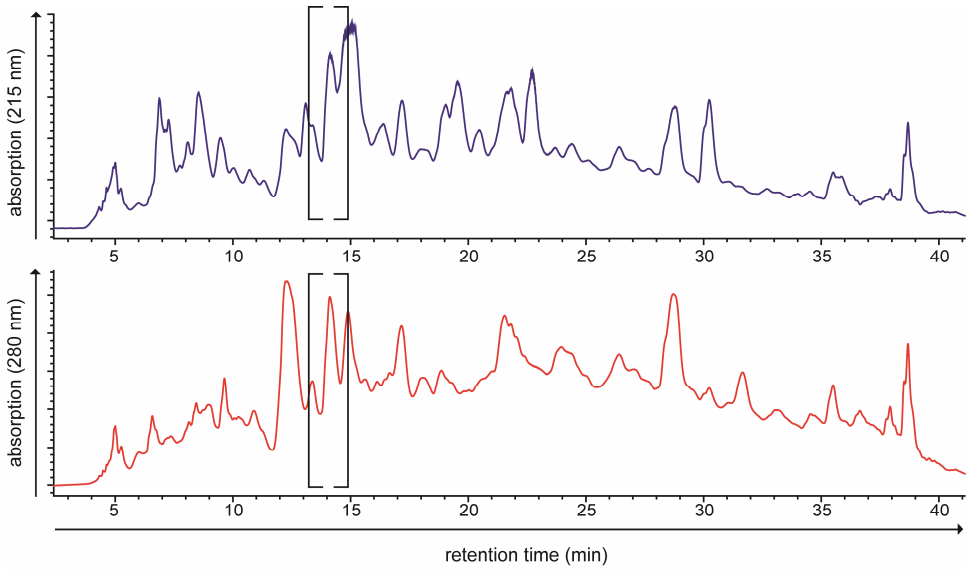



**Figure 5.13:** A preparative HPLC-profile of a XAD16 resin extraction of SCM-supernatants after 96 h growth of *S. erythraea*. The CAS-liquid assay positive fraction is indicated by brackets. The UV-absorption profiles clearly show the coelution of several compounds in the CAS-positive fraction due to high media complexity.

Bioinformatic analysis of the *etc* and the *nrps3* gene clusters led to the prediction of L-ornithine as a common building block putatively incorporated into both oligopeptides. In order to perform radiolabeling studies which target both compounds, a radio-LCMS-guided genome mining approach was applied by feeding the nonproteinogenic amino acid $^{14}$C-L-ornithine to cultures of *S. erythraea*.[225] The

experimental setup consisted of a HPLC-system coupled to a radioactivity flow-through detector and a Q-TOF mass spectrometer to directly link the radioactivity signal to a mass signal, leading to the identification of the target compound. $^{14}$C-L-Orn was added to the cultures prior to expression of the corresponding gene clusters (*etc*/*nrps3*) to ensure both the presence of the biosynthetic machinery and the incorporation of the radiolabel into the target compound (Chapter 4.7). Cultures of *S. erythraea*, supplemented with $^{14}$C-L-Orn, were grown for 96 h and extraction of the supernatant followed by radio-LCMS analysis revealed the radiolabeling of a compound with a measured m/z of 604.27 $[M+H]^+$ (Figure 5.14 A). The incorporation of radiolabeled L-Orn was determined to be 2% of the total amount of radioactivity fed to the cultures employing the SCM-medium. To reduce media complexity and to facilitate the isolation procedure, a radiolabeling experiment was carried out in iron-deficient M9 minimal medium. The incorporation of the radiolabel increased from 2% to 4% (Figure 5.14 B), whereas coeluting compounds were reduced as seen in the total ion chromatogram (TIC).

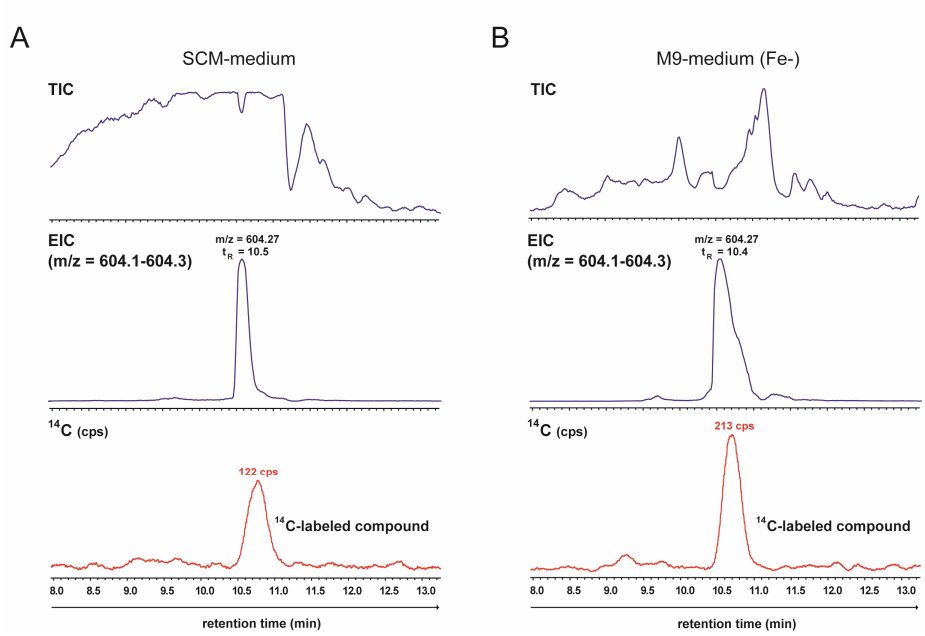
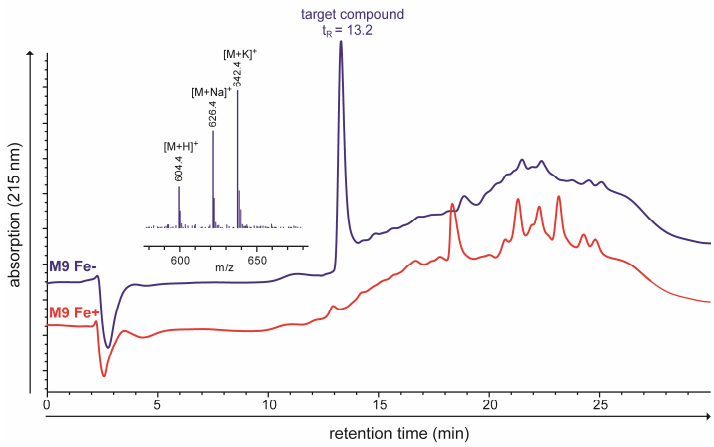



**Figure 5.14:** Radio-LCMS-profiles of radiolabeling experiments employing nonproteinogenic $^{14}$C-L-Orn. A) LCMS-trace of an extraction of SCM-medium. B) LCMS-trace of an extraction of iron-depleted M9 medium. In both cases the incorporation of the radiolabel occurred (red trace) and the compound eluting at $t_R$ = 10.5 displayed a discrete m/z = 604.27 ($[M+H]^+$) in the extracted ion chromatogram (EIC).

To isolate the siderophore in sufficient amounts for NMR-structure elucidation, a large-scale cultivation of *S. erythraea* in iron-deficient modified M9 medium was carried out, giving rise to siderophore production of 10.2 mg/L culture (Chapter 4.8.4). The physiological function of the siderophore for iron uptake was confirmed by comparing supernatant extractions of *S. erythraea* cultures grown in absence or presence of iron. The presence of iron in the medium completely suppressed siderophore production, whereas the absence of iron in the medium induced the production of the target compound which elutes at $t_R$ = 13.2 (Figure 5.15).




**Fugure 5.15:** LCMS-traces of XAD16 extractions of *S. erythraea* cultures grown in iron-depleted (blue trace) and iron-rich (red trace, 100 µM $FeCl_3$) M9 medium. The absence of iron clearly induces the production of the target compound which elutes at $t_R$ = 13.2 and shows the aforementioned m/z-ratios ($[M+H]^+$ = 604.4, $[M+Na]^+$ = 626.4, $[M+K]^+$ = 642.4), whereas the presence of iron completely supresses biosynthesis of the target compound.

### 5.2.3 NMR-Based Structure Elucidation of Erythrochelin

The structure of the target compound that was isolated from cultures of *S. erythraea* grown in M9 minimal medium was determined using NMR-methodology (Chapter 4.4.3) (Figure 5.16). The $^1H$ spectrum revealed the presence of four amide protons at 7.96, 7.74, 8.08, and 8.12 ppm (Figure S2). Four cross peaks were observed in the $^1H$-$^{15}N$ HSQC (heteronuclear single-quantum correlation spectroscopy) spectrum, which verified the presence of four amino acids in the sequence. The $^{15}N$ chemical shifts determined by cross peaks in the $^1H$-$^{15}N$ HSQC spectrum are listed in Table S5. TOCSY (total correlation spectroscopy) cross peaks confirmed the presence of three ornithines and one serine in the compound. Two strong singlets at 1.84 and 1.96 ppm for three and six protons, respectively, revealed the presence of three acetyl groups, of which two are attached to very similar amino acids in the sequence. The observed long-range $^1H$-$^{13}C$ correlations showed the two acetyl groups to be

connected to the δ-amino group of two δ-*N*-hydroxy ornithines, respectively, while the third one is attached to the α-amino group of one of the δ-*N*-acetyl-δ-*N*-hydroxyornithine (haOrn) residues, resulting in α-*N*-acetyl-δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine (ac-haOrn) (Figure 5.16). Three sequential NOE (nuclear Overhauser effect) contacts were observed, one revealing a connection between the terminal ac-haOrn$_1$ and the Ser$_2$, while the other two were for a sequential connection between a δ-*N*-hydroxyornithine and a δ-*N*-acetyl-δ-*N*-hydroxyornithine residue and its reverse, respectively. Such double sequential connections can only be established through a diketopiperazine (DKP) unit, which is composed of a hOrn and a haOrn moiety. Furthermore, a long-range $^{1}H–^{13}C$ correlation was detected between the carbonyl carbon of the serine and the δ-CH$_2$ of the hOrn, which constitutes the DKP. Analysis of the long-range connections enables the establishment of a structure for the tetrapeptide siderophore as shown in Figure 5.16. The assigned $^{1}H$, $^{13}C$ and $^{15}N$ chemical shifts are listed in Tables S3-S5. The observed NOE contacts and the long-range $^{1}H$-$^{13}C$ correlations verified the structure and are listed in Tables S6 and S7. Based on the results obtained by NMR the determined sequence for the peptide is ac-haOrn$_1$-Ser$_2$-cyclo(hOrn$_3$-haOrn$_4$). The corresponding DQF-COSY (double quantum filtering-correlation spectroscopy), $^{1}H–^{15}N$ HSQC, HMBC (heteronuclear multiple-bond correlation spectroscopy) and ROESY (rotating frame overhauser effect spectroscopy) spectra of erythrochelin are shown in Figures S3-S7.

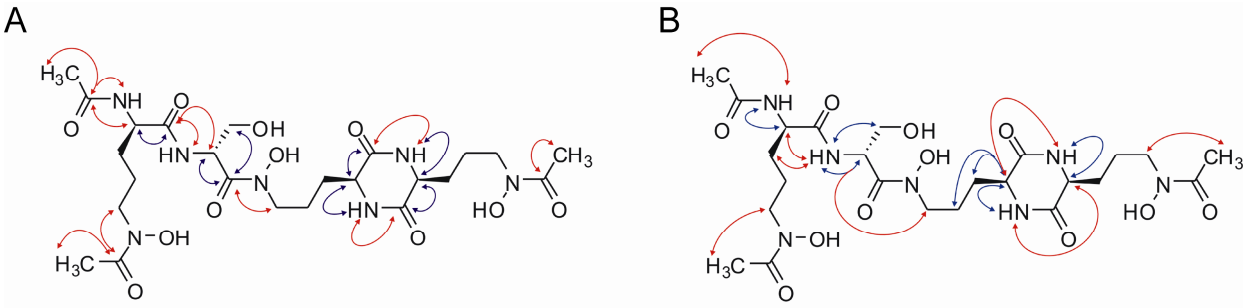



**Figure 5.16:** The structure of erythrochelin as determined by NMR. NMR-contacts are indicated by arrows. Blue arrows indicate intra-residue contacts; red arrows indicate long-range inter-residue contacts. A) Long-range $^{1}H–^{13}C$ correlations observed in DMSO (300K). B) NOE contacts observed in DMSO (300K). Sequential NOE contacts observed between hOrn$_3$ and haOrn$_4$ confirm the presence of a DKP-moiety.

### 5.2.4 Mass-Spectrometric- and Stereochemical Analysis of Erythrochelin

On the basis of the observed NMR-spectra the presence and connectivity of δ-*N*-acetyl-δ-*N*-hydroxyornithine, δ-*N*-hydroxyornithine and serine in the sequence was determined. Erythrochelin itself shows an exact m/z-ratio of 604.2938 ($[M+H]^+$, calculated 604.2937) and a molecular formula of $C_{24}H_{41}N_7O_{11}$ (Chapter 4.3.3). To confirm the structural assignment obtained by NMR, $MS^n$-fragmentation studies were conducted (Figure 5.17). An intense fragment with an m/z-ratio of 390.1979 ($[M+H]^+$, calculated 390.1983) corresponded to the *C*-terminal tripeptide comprised of serine and the DKP-moiety built up by hOrn and haOrn residues (Figure 5.17 A). The loss of the *N*-terminal serine residue gave rise to a dipeptidyl DKP-fragment with an m/z-ratio of 303.1662 ($[M+H]^+$, calculated 303.1663). This fragment was furthermore subjected to $MS^3$-fragmentation (Figure 5.17 B). The resulting fragments revealed the presence of hydroxylated and acetylated ornithine residues. In addition, an intense fragment with an m/z of 145.0869 ($[M+H]^+$, calculated 145.0971) was observed. This result gave strong evidence for the presence of the DKP-moiety since such fragmentation behaviour is characteristic for DKP-containing compounds and has been detected during fragmentation of an albonoursin intermediate (inset Figure 5.17 B).[226]

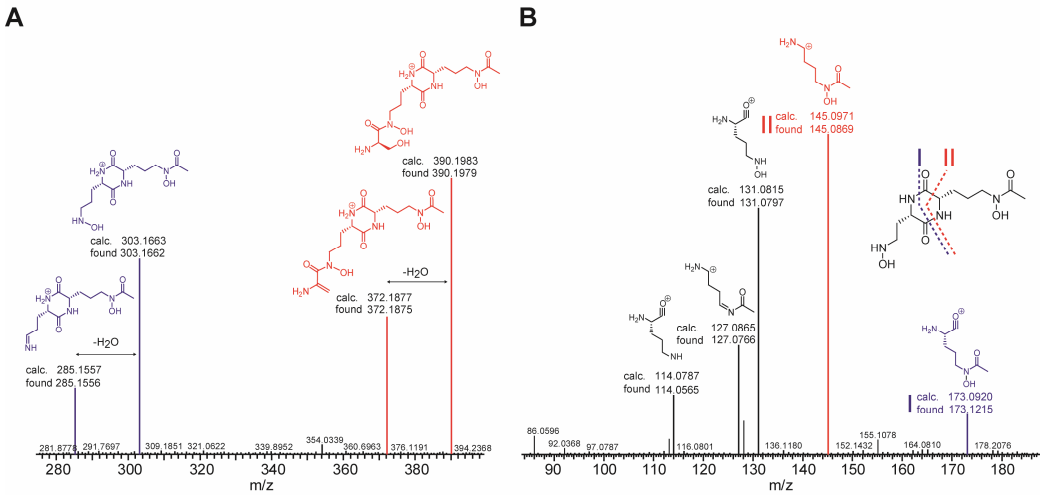



**Figure 5.17:** MS/MS-fragmentation studies of erythrochelin. A) $MS^2$-fragmentation of the title compound. The fragment displaying an m/z-ratio of 390.1979 ($[M+H]^+$, calculated 390.1983) corresponds to the *C*-terminal tripeptide (accentuated in red). Gas-phase fragmentation also results in the loss of the *N*-terminal serine-residue (accentuated in blue). The dipeptidyl fragment displays an m/z-ratio of 303.1662 ($[M+H]^+$, calculated 303.1663). B) $MS^3$-fragmentation pattern of the *C*-terminal DKP-moiety m/z = 303.1662 ($[M+H]^+$). Characteristic fragments resulting from DKP-fragmentation are accentuated in blue (Fragment I, m/z: 173.1215 ($[M+H]^+$, calculated 173.0920) or in red (Fragment II, m/z: 145.0869 ($[M+H]^+$, calculated 145.0971). Calculated and observed m/z-ratios for additional fragments are given.

Determination of overall stereochemistry of erythrochelin was carried out utilizing Marfey's reagent.[206] After the acid induced total hydrolysis of erythrochelin and subsequent lyophilization, a RP-LCMS analysis of the resulting products was carried out (Chapter 4.9.1). The HPLC-trace revealed the complete breakdown of erythrochelin and the loss of all δ-*N*-acetyl functionalities. Apart from serine and *N*-hydroxyornithine species no further compounds were detected. Based on these results the FDAA-derivatization procedure was carried out, employing the erythrochelin hydrolysate, the D/L-Ser and L-hOrn standards. Interestingly, the derivatization reaction of L-hOrn with FDAA in acetone led to the formation of a compound which did not show the expected mass of DNAP-L-hOrn. FT-MS data and $MS^2$-fragmentation proved the compound to be a nitrone species of L-hOrn. MS-data confirmed this result in both the derivatized standard as well as in the hydrolyzed erythrochelin sample. Therefore, the detected DNAP-hOrn compounds are designated DNAP-hOrn$^*$. These results were also obtained during the synthesis of ε-*N*-acetyl-ε-*N*-hydroxy-L-lysine and involve a nucleophilic attack of the hydroxylated side-chain amino group onto acetone.[227] After having established a methodology for the separation of the derivatized amino acid standards, the method was applied for the derivatized erythrochelin hydrolysate (Figure 5.18). The HPLC-chromatogram revealed the presence of D-configured serine and L-configured *N*-hydroxyornithine species. In addition, a third compound was detected showing the same MS-spectrum as DNAP-L-hOrn$^*$. This signal was assigned to DNAP-D-hOrn$^*$. Derivatization of the L-hOrn standard led solely to the formation of a monosubstituted DNAP-L-hOrn$^*$, excluding the formation of a δ-*N*-modified hOrn and clearly confirming the presence of D-hOrn in erythrochelin. Furthermore, the comparison of the UV/Vis-signal integrals, which show a ratio of 2:1 (L-hOrn:D-hOrn) gives strong evidence for the proposed structure and stereoconfiguration of erythrochelin. Coelution experiments, in which the derivatized erythrochelin hydrolysate was supplemented with synthetic standards, confirmed the previous observation (data not shown).

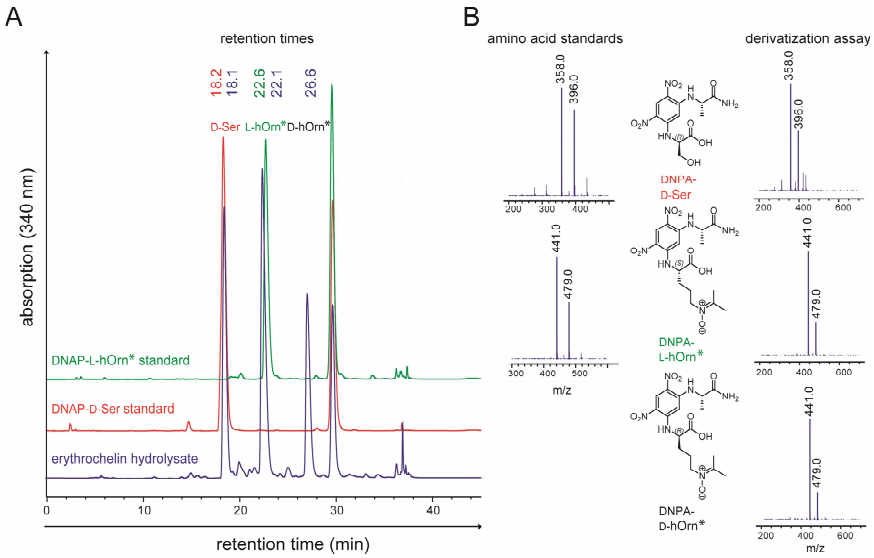



**Figure 5.18:** A) RP-LCMS traces of derivatized amino acid standards and the derivatization products of erythrochelin hydrolysis. The blue trace shows the derivatized erythrochelin hydrolysate. The signal observed at $t_R$ = 18.1 corresponds to derivatized DNPA-L-Ser as it coelutes with the DNPA-L-Ser standard (red trace). The signal observed at $t_R$ = 22.1 represents DNPA-L-hOrn* as it coelutes with the DNPA-L-hOrn* standard (green trace) and displays the same MS-spectrum. In addition to the amino acids D-Ser and L-hOrn a third building block ($t_R$ = 26.6) is present in erythrochelin showing the same MS-spectrum as DNPA-L-hOrn*. This signal was assigned to DNPA-D-hOrn*. B) The chemical structures and the observed MS-spectra of the derivatized amino acid standards and the derivatized products of erythrochelin hydrolysis.

In order to determine the connectivity of the amino acids as well as their stereoconfiguration resulting in the knowledge of the overall structure of erythrochelin, a partial hydrolysis-derivatization method was carried out (Chapter 4.9.2). Based on bioinformatic analysis of erythrochelin biosynthesis and the previous results obtained by FDAA-derivatization of the erythrochelin hydrolysate, it was proposed that the *C*-terminal DKP-moiety consists of two L-configured hOrn residues. To confirm this model, hydrolysis conditions were optimized towards the generation of the target fragment and the partially hydrolyzed *C*-terminal fragment was isolated. HRMS-analysis confirmed the identity of the isolated dipeptide (Figure 5.19 A). After isolation, the *C*-terminal fragment was hydrolyzed and derivatized. Analysis of the FDAA-derivatized amino acids confirmed the presence of solely L-configured hOrn residues (Figure 5.19 B). In addition to the MS-spectra and elution times compared to a synthetic DNAP-L-hOrn* standard, a coelution experiment was conducted, confirming the stereochemistry to be in full agreement with the proposed biosynthetic model. Taking all results together, the sequence of the tetrapeptide

siderophore erythrochelin was determined to be α-*N*-acetyl-δ-*N*-acetyl-δ-*N*-hydroxy-D-ornithine-D-serine-cyclo(δ-*N*-hydroxy-L-ornithine-δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine).

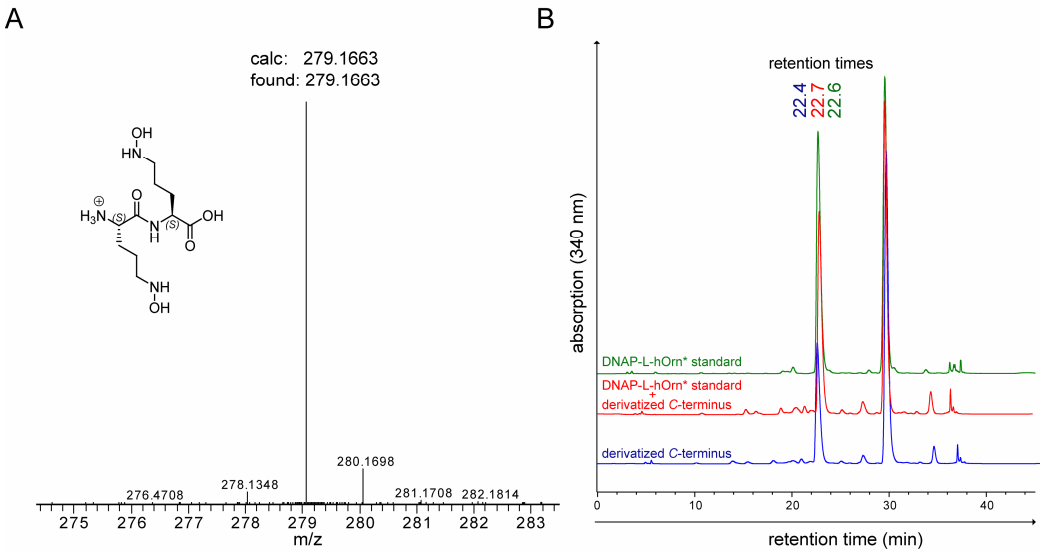



**Figure 5.19:** A) HRMS-analysis of the *C*-terminal fragment obtained after partial hydrolysis of erythrochelin. Calculated and observed m/z-ratios for the fragment are given. B) RP-LCMS traces of the hydrolyzed and derivatized *C*-terminal fragment and the DNAP-L-hOrn* standard. The hydrolysate alone is shown in blue whereas coelution with a synthetic DNAP-L-hOrn* standard is shown in red. The DNPA-L-hOrn* standard is shown in green. The signal observed at $t_R$ = 22.4 represents DNPA-L-hOrn* as it coelutes with the DNPA-L-hOrn* standard.

### 5.2.5 Physicochemical Properties of Erythrochelin

Siderophores represent iron-scavenging compounds (generally < 1 kDa) with a high affinity for ferric iron ($K_f$ = $10^{22}$ -$10^{49}$ $M^{-1}$) which form octahedral *ferri*-siderophore complexes.[146] To determine if erythrochelin inherits a physiological function as an iron-chelating compound, it was incubated with ferric iron in a 1:1-ratio and subjected to HRMS-analysis revealing a m/z of 657.2056 ($[M+H]^+$, calculated 657.2051) for the loaded *ferri*-erythrochelin. Analysis of *ferri*-erythrochelin via RP-LCMS revealed the lack of skimmer fragmentation as observed during analysis of *apo*-erythrochelin, indicating that the chelation of iron induces a structurally more rigid conformation which is less prone to fragmentation (Figure 5.20 A). MS-analysis furthermore confirmed the stoichiometry of the Fe(III):siderophore-complex to be 1:1, proving the presence of six Fe(III)-coordinating groups. As the binding of iron alters the spectrochemical properties of the siderophore by which the siderophore-types (*e.g.* hydroxamate, catecholate) can be distinguished corresponding spectra were recorded. UV/Vis-spectra of

*ferri*-erythrochelin compared to the unloaded *apo*-form show the typical absorption spectrum for hydroxamate-type siderophores ($\lambda_{max}$ = 440 nm), furthermore confirming the iron-chelating function of the product (Figure 5.20 B).

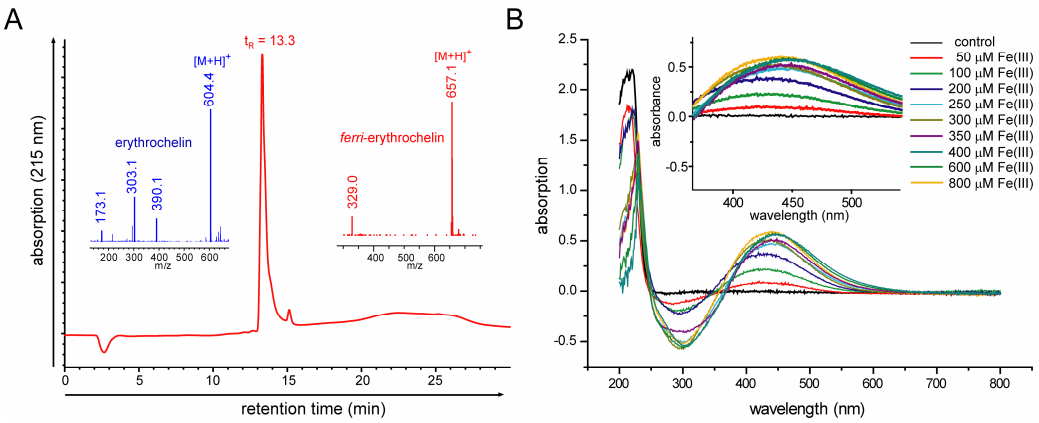



**Figure 5.20:** A) ESI-MS-analysis of *ferri*-erythrochelin $t_R$ = 13.3, m/z = 657.1 ($[M+H]^+$ , calculated 657.2). Skimmer fragmentation was completely abolished when analyzing *ferri*-erythrochelin (red), being indicative of a structurally rigid conformation induced upon iron-chelation. MS-analysis of erythrochelin is accentuated in blue. B) UV/Vis-absorption spectra of erythrochelin (400 µM) incubated with increasing concentrations of Fe(III). The absorption spectra are typical for hydroxamate-type siderophores with $\lambda_{max}$ = 440 nm. Incubation of erythrochelin with excess ferric iron exceeding 400 µM does not lead to a further increase in the absorption at λ = 440 nm, confirming the stochiometry of the Fe(III):siderophore-complex to be 1:1.

### 5.2.6 ACE-Inhibition Studies with Erythrochelin

The angiotensin-converting enzyme (ACE) is an exopeptidase, which is a central component of the renin-angiotensin system (RAS), found in higher eukaryotes, as it mediates extracellular volume and arterial vasoconstriction (Figure 5.21).[228] The RAS is initiated by the production of the protease renin in the kidneys in response to low blood pressure. Renin subsequently converts the 452 aa angiotensinogen to the decapeptide angiotensin I. ACE converts the decapeptide angiotensin I to the octapeptide angiotensin II through cleavage of the two *C*-terminal residues $His_9$ and $Leu_{10}$ (HL). The peptide hormone angiotensin II subsequently induces vasoconstriction and the rise of blood pressure.

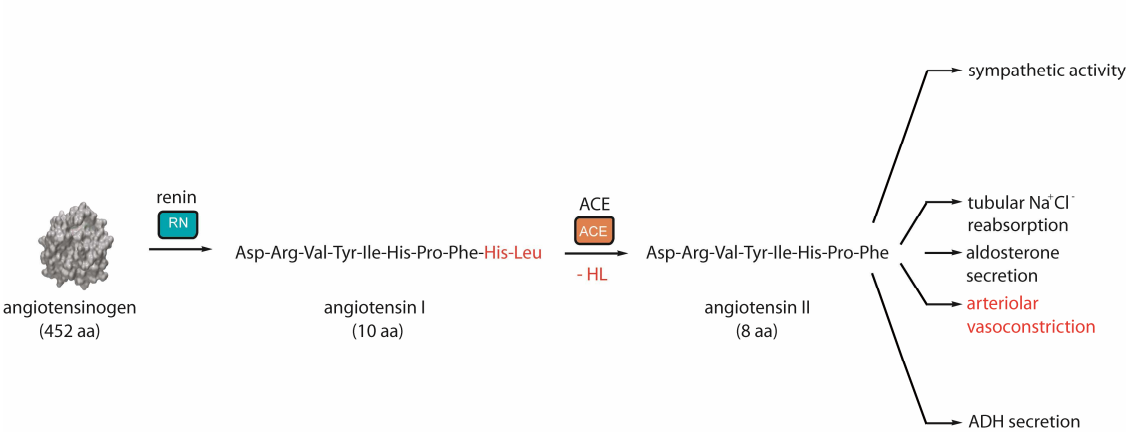



**Figure 5.21:** Overview of the renin-angiotensin system and the affected functions of the organism. The protease renin proteolytically cleaves the 452 aa angiotensinogen yielding in angiotensin I. Angiotensin I is subsequently converted into angiotensin II by the metalloprotease ACE. Angiotensin II then induces vasoconstriction and the rise of blood pressure (accentuated in red).

Erythrochelin shares a high degree of structural similarity to the siderophore foroxymithine isolated from cultures of *Streptomyces nitrosporeus*.[229-231] In contrast to erythrochelin the δ-amino groups of ac-$hOrn_1$ and $hOrn_4$ are formylated and all amino acids within the peptide chain show L-configuration. Foroxymithine has been reported as an inhibitor of ACE with an $IC_{50}$-value of 7 µg/mL.[229] Due to the structural similarities between foroxymithine and erythrochelin, the ACE-inhibitory effect of erythrochelin was investigated *in vitro* to evaluate the potential of the compound as a hypertension reducing agent (Chapter 4.10.2). In the assay the native substrate for ACE was substituted with the artificial tripeptide HHL (Figure 4.6). ACE-mediated proteolytic cleavage of the substrate leads to the release of hippuric acid, which can be detected via RP-LCMS.[208] A standard curve for the determination of hippuric acid release was obtained by using stock solutions and measuring the absorption at 228 nm (Figure 5.22 A). Assay solutions were incubated with increasing amounts of erythrochelin (5 - 250 µM) and *ferri*-erythrochelin (5 - 250 µM) and analyzed via RP-LCMS. Throughout the assays no inhibition of ACE could be observed as the release of hippuric acid corresponded to the control lacking erythrochelin (Figure 5.22 B). The lack of ACE inhibition confirmed the structural features of foroxymithine to be essential for interaction with the metalloprotease.

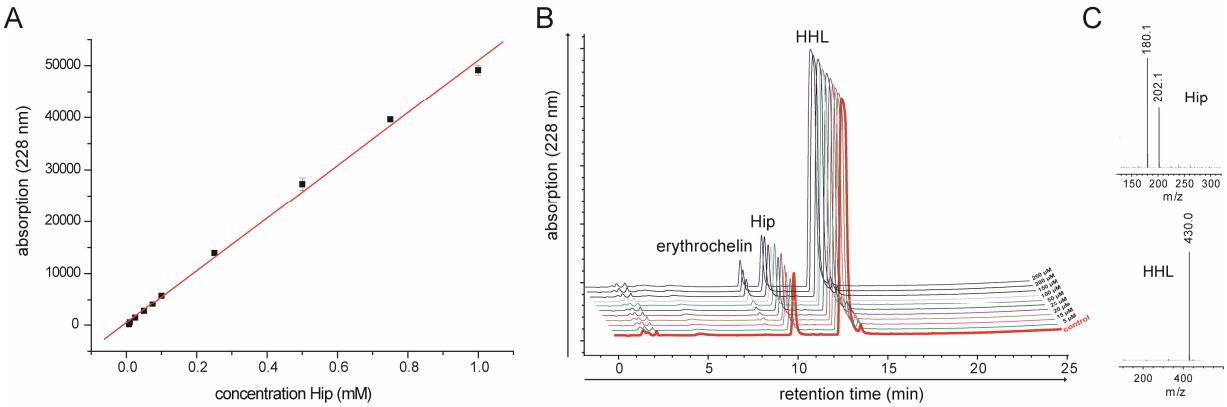



**Figure 5.22:** A) The hippuric acid (Hip) standard curve for the determination of HHL cleavage by the ACE. B) HPLC-traces of the *in vitro* ACE-inhibition assays with increasing concentrations of erythrochelin (5 – 250 µM) ($t_R$ = 8.7). The release of hippuric acid ($t_R$ = 10.1) from the artificial substrate HHL ($t_R$ = 12.9) results from proteolytic activity of the ACE. The control lacking erythrochelin is accentuated in red. As increasing concentrations of the potential inhibitor do not decrease the degree of hippuric acid release, it is proposed that the structure of erythrochelin impedes inhibition of the ACE. C) The observed ESI-MS-spectra of hippuric acid (Hip) and HHL.

### 5.3 Biosynthesis of the δ-*N*-Hydroxy-L-Ornithine Residue of Erythrochelin

Erythrochelin is a hydroxamate-type tetrapeptide siderophore assembled by a nonribosomal machinery capable of chelating iron with high affinity. It contains three ornithine residues of which two are δ-*N* acetylated and δ-*N* hydroxylated and one is solely δ-*N* hydroxylated. These groups contribute to the iron-chelating properties of erythrochelin and are involved in the octahedral coordination of ferric iron. A-domain specificity prediction resulted in δ-*N*-hydroxy-L-ornithine (hOrn) being the substrate for the third adenylation domain found in the NRPS EtcD (Table 5.1). In addition, bioinformatic analysis of the *etc* gene cluster revealed the gene *etcB* to encode a FAD-dependent monooxygenase putatively involved in the generation of hOrn from L-ornithine (Figure 5.11). It shows 64% sequential identity and 78% similarity to the characterized FAD-dependent monooxygenase CchB from *S. coelicolor,* involved in coelichelin biosynthesis.[98] To gain further insights into the biosynthesis of erythrochelin, EtcB was recombinantly produced and isolated. Hydroxylation assays employing the recombinant enzyme, the cosubstrates and L-Orn gave rise to the building block δ-*N*-hydroxy-L-ornithine, which represents a branching point in erythrochelin biosynthesis (Chapter 5.4). The substrate specificity of EtcB and the kinetic parameters of the hydroxylation reaction were determined, proving EtcB to be a typical member of the FAD-dependent monooxygenase family.

#### 5.3.1 Recombinant Expression and Isolation of EtcB as an Active *Holo*-Protein

*EtcB* was amplified from the chromosomal DNA of *Saccharopolyspora erythraea* NRRL 23338 and cloned into the pCB28a(+) vector. It was heterologously expressed in *E. coli* BL21(DE3) cells (Chapter 4.1.5) and isolated as a *C*-terminally $His_6$-tagged *holo*-protein (Figure 5.23 A) as described in the Methods section (Chapter 4.2.1.1). Final protein yield per liter culture after concentration and dialysis was 6.6 mg. UV/Vis-spectroscopic analysis confirmed the recombinant protein to be loaded with the proposed cofactor FAD (Figure 5.23 B).

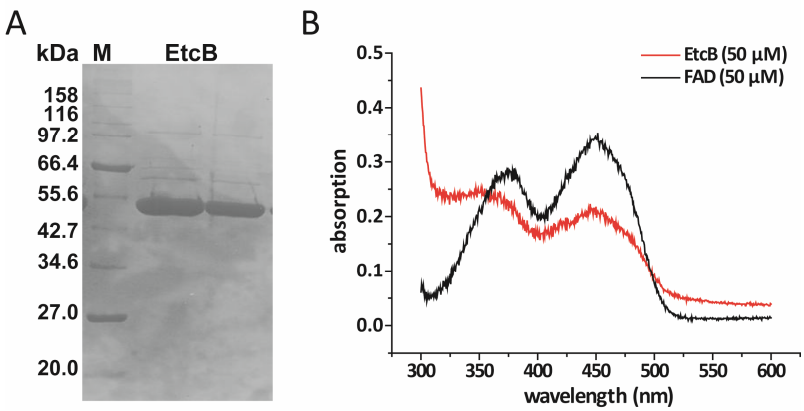



**Figure 5.23:** A) SDS-PAGE of purified recombinant EtcB (51.3 kDa). Protein Marker (M) was Broad Range Protein Marker P7702 (NEB). B) UV/Vis-spectrum of EtcB (red) compared to the absorption spectrum of FAD alone (black). Absorption maxima at 370 nm and 450 nm indicate the presence of an enzyme-bound FAD-cofactor.

### 5.3.2 Biochemical Characterization of EtcB

Sequence analysis and alignments showed that EtcB is a FAD-dependent monooxygenase, which requires NADPH as the reducing cosubstrate and molecular oxygen to mediate δ-*N*-hydroxylation of L-ornithine. To investigate if EtcB is catalyzing δ-*N*-hydroxylation of L-ornithine, it was incubated with the substrate L-Orn and the corresponding cosubstrates (Chapter 4.6.4). RP-LCMS analysis of the assays after 4 h revealed 50% conversion of L-ornithine ($t_R$ = 4.8, m/z = 133.0 $[M+H]^+$ observed, m/z = 133.1 $[M+H]^+$ calculated) to δ-*N*-hydroxy-L-ornithine (hOrn) ($t_R$ = 13.7, m/z = 149.0 $[M+H]^+$ observed, m/z = 149.1 $[M+H]^+$ calculated) in the presence of the enzyme (Figure 5.24). HRMS-analysis confirmed the product to be a hydroxylated ornithine species (m/z = 149.0922 $[M+H]^+$ observed, m/z = 149.0921 $[M+H]^+$ calculated). Substrate turnover could not be detected in the absence of EtcB or NADPH. In addition, the utilization of NADH as the reducing cosubstrate did not lead to the formation of hOrn. The identity of the product was confirmed by comparison with synthetically obtained L-hOrn.[207] The product of the enzymatic assay and the synthetic standard displayed the same retention times and MS-spectra, when subjected to RP-LCMS analysis.

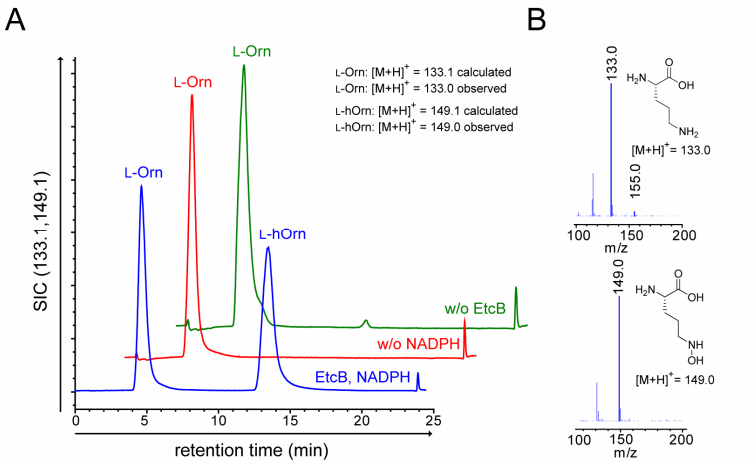



**Figure 5.24:** δ-*N*-hydroxylation of L-ornithine catalyzed by EtcB. A) LCMS-traces of the hydroxylation assays are shown as selected ion chromatograms (SIC). Incubation of the substrate L-Orn ($t_R$ = 4.8) with EtcB and the cosubstrates gives rise to the product L-hOrn ($t_R$ = 13.7). The control lacking the reducing cosubstrate NADPH is shown in red and the control lacking EtcB is shown in green. The calculated and observed m/z-ratios are given in the inset. B) The chemical structures and the observed ESI-MS-spectra of the substrate L-Orn (m/z = 133.0 $[M+H]^+$) and the product L-hOrn (m/z = 149.0 $[M+H]^+$).

To evaluate the substrate specificity of EtcB, the recombinant enzyme was incubated with a set of alternative substrates representing the different classes of amino acids. In addition, the D-isomer of L-Orn was employed as well as α-*N*-acetyl-L-ornithine (L-acOrn). Assays were carried out according to the conditions described above (Chapter 4.6.4) and were analyzed via RP-LCMS. Hydroxylation was exclusively limited to L-ornithine as shown in Table 5.3.

**Table 5.3:** Overview of the substrates evaluated for EtcB-mediated hydroxylation. The corresponding m/z-ratios are given for each substrate and hydroxylated substrate.

| substrate | m/z $[M+H]^+$ substrate | m/z $[M+H]^+$ hydroxylated substrate | m/z $[M+H]^+$ observed | hydroxylation |
|---|---|---|---|---|
| L-Orn | 133.1 | 149.1 | 149.0 | ✓ |
| D-Orn | 133.1 | 149.1 | 133.0 | X |
| L-acOrn | 175.2 | 191.2 | 175.0 | X |
| L-Lys | 147.1 | 163.1 | 147.0 | X |
| L-Val | 118.1 | 134.1 | 118.0 | X |
| L- Asp | 134.1 | 150.1 | 134.0 | X |
| L-Asn | 133.1 | 149.1 | 133.1 | X |
| L-Glu | 148.0 | 164.0 | 148.1 | X |
| L-Gln | 147.1 | 163.1 | 147.2 | X |
| L-Arg | 175.2 | 191.2 | 175.3 | X |

Kinetic parameters for EtcB-catalyzed δ-*N*-hydroxylation of L-Orn were determined spectrophotometrically by measurement of the initial rates of the assay employing different substrate concentrations. The calculation of kinetic parameters was carried out with the Enzyme Kinetics Module for SigmaPlot 8.0 (Chapter 4.6.4) using Michalis-Menten and Lineweaver-Burk equations (Figure 5.25).

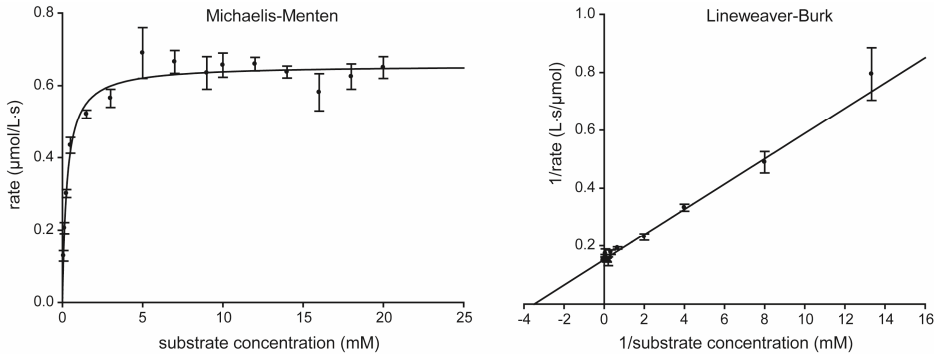



**Figure 5.25:** Michaelis-Menten and Lineweaver-Burk plots for kinetic parameter determination of EtcB.

The kinetic parameters for EtcB-mediated δ-*N*-hydroxylation of L-Orn were calculated to an apparent $K_M$ of 0.286 ± 0.035 mM and $k_{cat}$ = 19.6 ± 0.03 $min^{-1}$, leading to a catalytic efficiency of $k_{cat}/K_M$ = 68.5 $min^{-1} \cdot mM^{-1}$. Taking the results together it can be stated that EtcB can provide the L-hOrn building block required for NRPS-based erythrochelin assembly.

### 5.4 Acetylation of δ-*N*-Hydroxy-L-Ornithine During Erythrochelin Biosynthesis

Hydroxamate-type siderophores generally contain unusual *N*-acyl, *N*-hydroxy- or *N*-acyl-*N*-hydroxy amino acids in the oligopeptide chain, which are essential for the coordination of iron. Although the hydroxylation of L-lysine or L-ornithine residues has been extensively characterized, the mechanism and timing of side-chain acylation remains to be elucidated.[98] Erythrochelin itself contains two δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine (L-haOrn) residues, which contribute to the iron-chelating properties of the compound. Bioinformatic analysis identified two enzymes, Sace_1309 and Mcd presumably involved inL-haOrn biosynthesis. In order to investigate if both enzymes participate in the biosynthesis of L-haOrn, they were produced in a recombinant manner and subjected to biochemical studies *in vitro*. Hydroxylation assays employing recombinant Sace_1309, the cosubstrates and L-Orn gave rise to δ-*N*-hydroxy-L-ornithine and confirmed the enzyme to be involved in building block assembly. Acetylation assays were carried out by incubating L-hOrn with recombinant Mcd and malonyl-CoA as the acetyl-group donor. Substrate turnover was increased by substituting malonyl-CoA with acetyl-CoA,

bypassing the decarboxylation reaction which represents the rate-limiting step. Consecutive enzymatic synthesis of L-haOrn was accomplished in coupled assays employing both an L-ornithine hydroxylase and Mcd. In summary, a biosynthetic route for the generation of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine starting from L-ornithine has been established *in vitro* by tandem action of a FAD-dependent monooxygenase (EtcB or Sace_1309) and the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd.

### 5.4.1 Bioinformatic Identification of the Bifunctional Enzyme Mcd

Erythrochelin assembly requires the proliferation of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine, which is incorporated into the tetrapeptide at positions 1 and 4. Usually, enzymes involved in the modification or synthesis of specific building blocks for natural product assembly are integral components of the corresponding gene clusters. In the case of coelichelin the enzymes responsible for the synthesis of the δ-*N*-formyl-δ-*N*-hydroxy-L-ornithine residues are encoded within the *cch* biosynthetic gene cluster.[38,98] In contrast, the *etc* gene cluster does not house a gene predicted to encode an acetyltransferase capable of catalyzing the acetylation of δ-*N*-hydroxy-L-ornithine. In order to identify a putative acetyltransferase within the genome of *S. erythraea* (GenBank accession number: AM420293), a genome-genome BLAST analysis was carried out with the query sequence being the genome of the aerobactin producing strain *Shigella flexneri* 5 str. 8401 (GenBank accession number: CP000266). Aerobactin itself represents a NIS-derived hydroxamate-type siderophore in which the ε-*N*-hydroxy functionalities of L-lysine are acetylated.[232] It was proposed that the putative *N*-acetyltransferase IucB is responsible for *N*-acetyl transfer as it was shown to be required for aerobactin production.[233] Screening of the *S. erythraea* genome for *iucB*-type genes using ACT revealed the only significant hit to be *mcd* (*sace_1304*), encoded within the *nrps1* gene cluster. This gene displayed 33% sequential identity and 49% sequential similarity to *iucB* (Figure 5.26). Due to multiple sequence alignments it was proposed that the *C*-terminal region of Mcd represents the *N*-acetyltransferase- and malony-CoA decarboxylase domain similar to CurA, involved in curacin A (*Lyngbya majuscula*) biosynthesis.[234] Intriguingly, Mcd has already been characterized and shown to catalyze the biotin-independent decarboxylation of malonyl-CoA.[235-236] Based on the two functionalities of Mcd it was postulated that Mcd catalyzes the decarboxylation of malony-CoA to acetyl-CoA, which is subsequently transferred onto the δ-amino group of δ-*N*-hydroxy-L-ornithine to generate δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine. In addition to *mcd*, a second gene is located within both biosynthetic gene clusters (Figure 5.26). *Sace_1309* and *iucD* both share a high degree of sequential identity and are annotated as FAD-dependent monooxygenases, involved in siderophore biosynthesis. IucD is proposed to catalyze the hydroxylation of L-lysine during aerobactin biosynthesis

(Figure 1.19), whereas a specific function for Sace_1309 could not be predicted as the product of *nrps1* is unknown and a frameshift mutation has been allocated in *nrps1-1*. [215,237] Sace_1309 itself shares 58% sequential identity and 70% sequential similarity to CchB, the hydroxylase involved in coelichelin biosynthesis.[98] To obtain further insights into the biosynthesis of erythrochelin and to establish a biosynthetic route for the generation of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine, Sace_1309 and Mcd were subjected to biochemical studies *in vitro*.

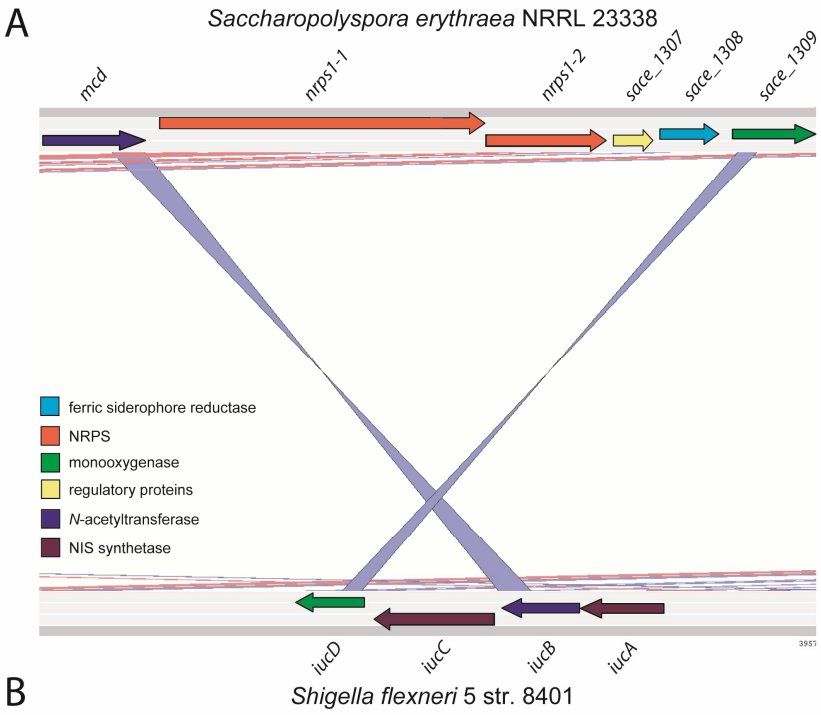



**Figure 5.26:** Graphical result of the genome-genome BLAST analysis with the Artemis Comparison Tool (ACT). Putative functions of the proteins encoded within the clusters are based on BLAST-analysis and are given in the figure. A) Organization of the *nrps1* gene cluster (*Saccharopolyspora erythraea* NRRL 23338). B) Organization of the aerobactin gene cluster (*Shigella flexneri* 5 str. 8401). The genes *iucB* and *iucD* share a high degree of sequential identity to *mcd* and *sace_1309*, encoding a putative acetyltransferase and a FAD-dependent monooxygenase. Regions of synteny are accentuated in light blue.

### 5.4.2 Recombinant Expression and Isolation of Sace_1309 as an Active *Holo*-Protein

*Sace_1309* was amplified from the chromosomal DNA of *Saccharopolyspora erythraea* NRRL 23338 and cloned into the pCB28a(+) vector. It was heterologously expressed in *E. coli* BL21(DE3) cells (Chapter 4.1.5) and isolated as a *C*-terminally $His_6$-tagged *holo*-protein (Figure 5.27 A) as described in the Methods section (Chapter 4.2.1.1). Final protein yield per liter culture after concentration and dialysis was 8.6 mg. UV/Vis-spectroscopic analysis confirmed the recombinant protein to be loaded with the proposed cofactor FAD (Figure 5.27 B).

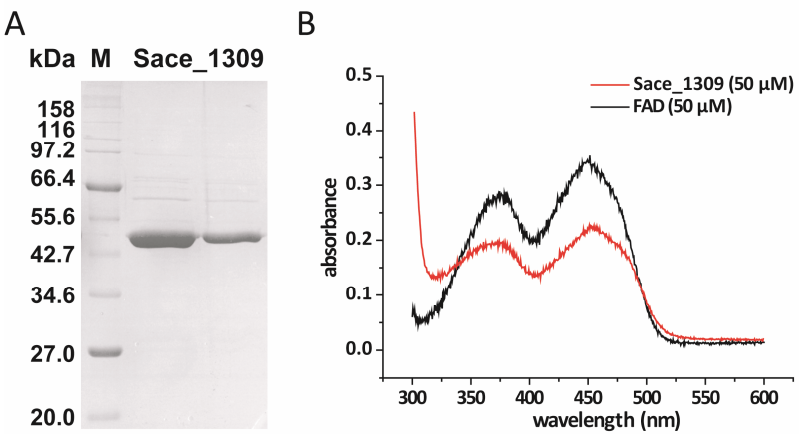



**Figure 5.27:** A) SDS-PAGE of purified recombinant Sace_1309 (40.9 kDa). Protein Marker (M) was Broad Range Protein Marker P7702 (NEB). B) UV/Vis-spectrum of Sace_1309 (red) compared to the absorption spectrum of FAD alone (black). Absorption maxima at 370 nm and 450 nm indicate the presence of an enzyme-bound FAD-cofactor.

### 5.4.3 Biochemical Characterization of Sace_1309

Sequence analysis and alignments revealed that Sace_1309 is a FAD-dependent monooxygenase requiring NADPH as the reducing cosubstrate and molecular oxygen to mediate δ-*N*-hydroxylation of L-ornithine. To investigate if Sace_1309 is catalyzing δ-*N*-hydroxylation of L-ornithine in analogy to EtcB (Chapter 5.3.2) it was incubated with the substrate L-Orn and the corresponding cosubstrates (Chapter 4.6.4). RP-LCMS analysis of the assays after 4 h revealed 35% conversion of L-ornithine ($t_R$ = 5.9, m/z = 133.0 $[M+H]^+$ observed, m/z = 133.1 $[M+H]^+$ calculated) to δ-*N*-hydroxy-L-ornithine (hOrn) ($t_R$ = 14.0, m/z = 149.0 $[M+H]^+$ observed, m/z = 149.1 $[M+H]^+$ calculated) in the presence of the enzyme (Figure 5.28). Substrate turnover could not be detected in the absence of Sace_1309 or NADPH or when employing NADH as the reducing cosubstrate.

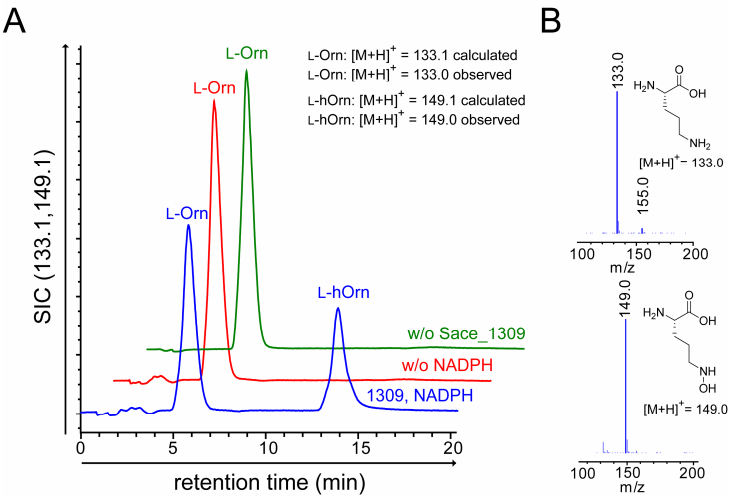



**Figure 5.28:** δ-*N*-hydroxylation of L-ornithine catalyzed by Sace_1309. A) LCMS-traces of the hydroxylation assays are shown as selected ion chromatograms (SIC). Incubation of the substrate L-Orn ($t_R$ = 5.9) with Sace_1309 and the cosubstrates gives rise to the product L-hOrn ($t_R$ = 14.0). The control lacking the reducing cosubstrate NADPH is shown in red and the control lacking Sace_1309 is shown in green. The calculated and observed m/z-ratios are given in the inset. B) The chemical structures and the observed ESI-MS-spectra of the substrate L-Orn (m/z = 133.0 $[M+H]^+$) and the product L-hOrn (m/z = 149.0 $[M+H]^+$).

Substrate specificity of Sace_1309 was evaluated by incubating the recombinant enzyme with a set of alternative substrates representing the different classes of amino acids. In addition, the D-isomer of L-Orn was employed as well as α-*N*-acetyl-L-ornithine (L-acOrn). Assays were carried out according to the conditions described above (Chapter 4.6.4) and were analyzed via RP-LCMS. In analogy to the results obtained during the biochemical characterization of EtcB (Chapter 5.3.2), hydroxylation was exclusively limited to L-ornithine and proved Sace_1309 to be responsible for the generation of L-hOrn.

### 5.4.4 Recombinant Expression and Isolation of Mcd

*Mcd* was amplified from the chromosomal DNA of *Saccharopolyspora erythraea* NRRL 23338 and cloned into the pET28a(+) vector. It was heterologously expressed in *E. coli* BL21(DE3) cells (Chapter 4.1.5) and isolated as a *C*-terminally $His_6$-tagged protein (Figure 5.29) as described in the Methods section (Chapter 4.2.1.1). Final protein yield per liter culture after concentration and dialysis was 9.6 mg.

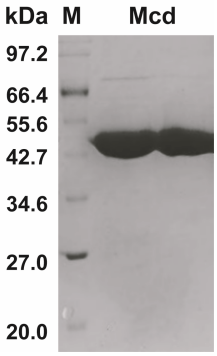



**Figure 5.29:** SDS-PAGE of purified recombinant Mcd (48.4 kDa). Protein Marker (M) was Broad Range Protein Marker P7702 (NEB).

### 5.4.5 Biochemical Characterization of the Bifunctional Enzyme Mcd

Bioinformatic analysis of Mcd revealed that the enzyme contains two putative functional domains: one domain responsible for the decarboxylation of malonyl-CoA (m-CoA) to afford acetyl-CoA (ac-CoA) and an acetyltransferase domain to catalyze the acetylation of L-ornithine or δ-*N*-hydroxy-L-ornithine (Chapter 5.4.1). To investigate if Mcd catalyzes the decarboxylation of malonyl-CoA and subsequent acetyltransfer, it was incubated with the substrates L-Orn and malonyl-CoA (Chapter 4.6.5). RP-LCMS analysis of the assay after 8 h showed that L-Orn is not the cognate substrate for Mcd-mediated acetylation as no substrate turnover was observed. In contrast, incubation of Mcd with L-hOrn and malonyl-CoA as the acetyl-donor afforded acetylation of the substrate. RP-LCMS analysis of the assays after 4 h revealed 38% conversion of L-hOrn ($t_R$ = 12.1, m/z = 149.0 $[M+H]^+$ observed, m/z = 149.1 $[M+H]^+$ calculated) to δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine (haOrn) ($t_R$ = 13.5, m/z = 191.0 $[M+H]^+$ observed, m/z = 191.1 $[M+H]^+$ calculated) in the presence of the enzyme (Figure 5.30). Substrate conversion was increased to 79% after 8 h. Substrate turnover could not be detected in the absence of Mcd or malonyl-CoA. HRMS-analysis of the product confirmed it to be the acetylated hydroxyornithine species (m/z = 191.1028 $[M+H]^+$ observed, m/z = 191.1026 $[M+H]^+$ calculated) (Figure 5.30 C). Future work will focus on proving acetylation of the δ-hydroxamino group by chemical synthesis of a L-haOrn standard, but recent genetic studies confirmed the results presented herein (Chapter 6.3.2.2).[238]

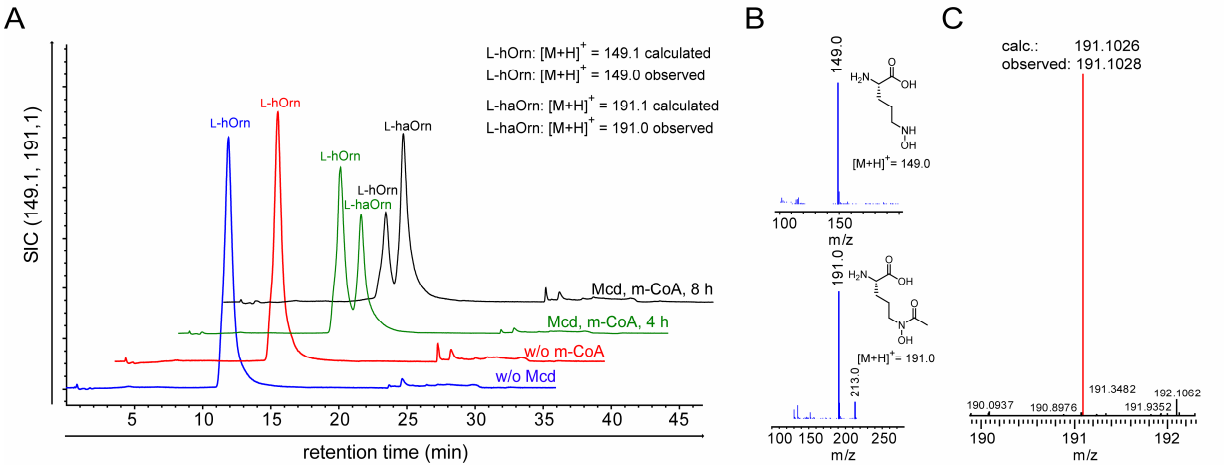



**Figure 5.30:** Mcd-mediated acetylation of L-hOrn. A) LCMS-traces of the acetylation assays are shown as selected ion chromatograms (SIC). Incubation of the substrate L-hOrn ($t_R$ = 12.1) with Mcd and malonyl-CoA gives rise to the product L-haOrn ($t_R$ = 13.5, green and black trace). The control lacking Mcd is shown in blue and the control lacking malonyl-CoA is shown in red. The calculated and observed m/z-ratios are given in the inset. B) The chemical structures and the observed ESI-MS-spectra of the substrate L-hOrn (m/z = 149.0 $[M+H]^+$) and the product L-haOrn (m/z = 191.0 $[M+H]^+$). C) HRMS-analysis of L-haOrn (m/z = 191.1028 $[M+H]^+$ observed, m/z = 191.1026 $[M+H]^+$ calculated).

Acetylation of the cognate substrate L-hOrn by Mcd requires the decarboxylation of malonyl-CoA, which precedes the acetyltransfer. In addition, substrate channelling of the resulting acetyl-CoA from the decarboxylase domain to the acetyltransferase domain has to occur simultaneously to the binding of L-hOrn.[234] In order to circumvent the decarboxylation reaction and to enable direct acetyltransfer, acetyl-CoA was employed as an alternative acetyl-group donor. Assays were carried out and analyzed by RP-LCMS as described earlier (Chapter 4.6.5). Analysis of the assays revealed total substrate turnover after 4 h (Figure 5.31 A). Substrate turnover could not be detected in the absence of Mcd or acetyl-CoA. This result demonstrated that Mcd is quantitatively acetylating L-hOrn in the presence of the acetyl-group donor acetyl-CoA.

As malonyl-decarboxylation and subsequent substrate channelling appeared to be the rate-limiting step of L-hOrn acetylation, a time course experiment was carried out to investigate substrate conversion depending on the acetyl-group donor. As substrate conversion occurred rapidly when utilizing acetyl-CoA, the concentration of Mcd was reduced to 1 µM and samples were taken each 30 min and analyzed via RP-LCMS. Determination of substrate conversion was based on integration of MS-signals in the selected ion chromatograms (SIC). Analysis of the assays revealed complete substrate turnover in the presence of acetyl-CoA after 90 min (Figure 5.31 B). In contrast, the utilization of malonyl-CoA as the

acetyl-donor resulted in only 35% substrate conversion after 240 min. These results indicate that malonyl-CoA decarboxylation is the rate-limiting step which can be bypassed by directly supplying acetyl-CoA as the cosubstrate for Mcd.

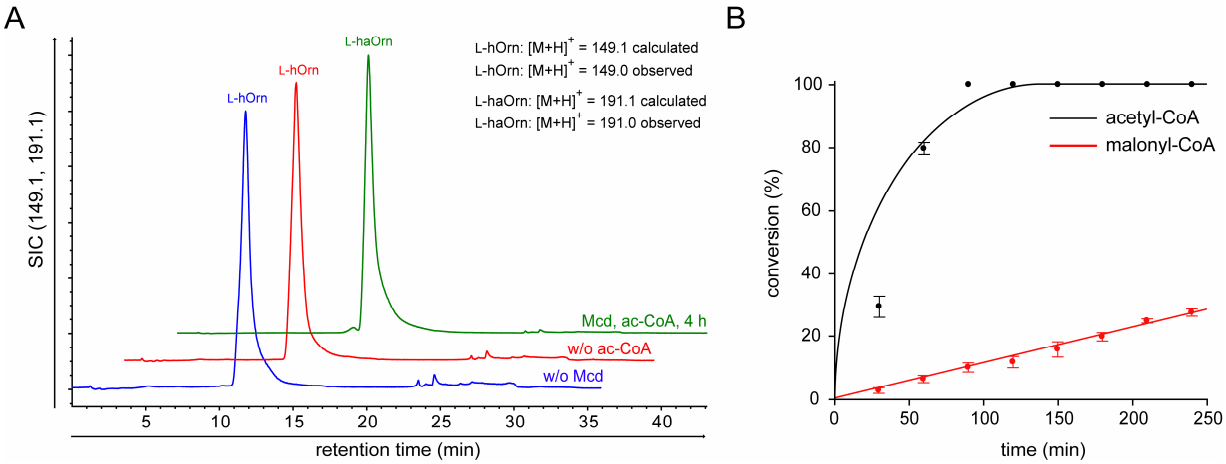



**Figure 5.31:** Mcd-mediated acetylation of L-hOrn. A) LCMS-traces of the acetylation assays are shown as selected ion chromatograms (SIC). Incubation of the substrate L-hOrn ($t_R$ = 12.1) with Mcd and acetyl-CoA gives rise to the product L-haOrn ($t_R$ = 13.2). The control lacking Mcd is shown in blue and the control lacking acetyl-CoA is shown in red. The calculated and observed m/z-ratios are given in the inset. B) Time dependent conversion of L-hOrn to L-haOrn in the presence of either malonyl-CoA (red) or acetyl-CoA (black).

Substrate specificity of Mcd was evaluated by incubating the recombinant enzyme with a set of alternative substrates representing the different classes of amino acids. Assays were carried out according to the conditions described above by using both malony-CoA and acetyl-CoA as the acetyl-group donor (Chapter 4.6.5) and were analyzed via RP-LCMS. Acetylation was exclusively limited to L-hOrn as shown in Table 5.4.

**Table 5.4:** Overview of the substrates evaluated for Mcd-mediated acetylation. The corresponding m/z-ratios are given for each substrate and acetylated substrate.

| substrate | m/z $[M+H]^+$ substrate | m/z $[M+H]^+$ acetylated substrate | m/z $[M+H]^+$ observed | acetylation |
|---|---|---|---|---|
| L-hOrn | 149.1 | 191.1 | 191.1 | ✓ |
| L-Orn | 133.1 | 175.1 | 133.0 | X |
| D-Orn | 133.1 | 175.1 | 133.0 | X |
| L-acOrn | 175.2 | 217.1 | 175.0 | X |
| L-Lys | 147.1 | 189.1 | 147.0 | X |
| L-Val | 118.1 | 160.1 | 118.0 | X |
| L- Asp | 134.1 | 176.1 | 134.0 | X |
| L-Asn | 133.1 | 175.1 | 133.1 | X |
| L-Glu | 148.0 | 190.1 | 148.0 | X |
| L-Gln | 147.1 | 189.1 | 147.2 | X |
| L-Arg | 175.2 | 217.1 | 175.3 | X |

### 5.4.6 Consecutive Enzymatic Synthesis of δ-*N*-Acetyl-δ-*N*-Hydroxy-L-Ornithine

Biosynthesis of the building block δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine requires the tandem action of two distinct enzymes. In the biosynthetic scheme based on the results obtained it is postulated that the first step is L-ornithine hydroxylation catalyzed by the FAD-dependent monooxygenases EtcB or Sace_1309. This hydroxylated ornithine species is subsequently acetylated by Mcd, which utilizes either malonyl-CoA or acetyl-CoA as the acetyl-group donor. As acetylation of L-Orn could not be observed *in vitro*, the alternative biosynthetic scheme in which acetylation of ornithine precedes hydroxylation was ruled out. To establish a biosynthetic route for the generation of L-haOrn, a coupled assay employing both the FAD-dependent monooxygenase and the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd was carried out (Chapter 4.6.6). In these assays, both malonyl-CoA and acetyl-CoA served as cosubstrates for Mcd-mediated acetyl-transfer. RP-LCMS analysis of the assay with EtcB, Mcd and malonyl-CoA revealed 56% conversion of L-Orn ($t_R$ = 4.5, m/z = 133.0 $[M+H]^+$ observed, m/z = 133.1 $[M+H]^+$ calculated) to δ-*N*-hydroxy-L-ornithine (hOrn) ($t_R$ = 12.2, m/z = 149.0 $[M+H]^+$ observed, m/z = 149.1 $[M+H]^+$ calculated) and 32% conversion to δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine (haOrn) ($t_R$ = 13.1, m/z = 191.0 $[M+H]^+$ observed, m/z = 191.1 $[M+H]^+$ calculated) after 8 h (Figure 5.32). The same assay conducted with acetyl-CoA as acetyl-donor gave rise to an almost quantitative conversion of L-Orn to L-haOrn and confirmed the results obtained during the biochemical characterization of Mcd (Chapter 5.4.5). Similar results were obtained when substituting the hydroxylase EtcB with Sace_1309, encoded in the *nrps1* gene cluster. In summary, consecutive enzymatic synthesis of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine is achieved by tandem

action of a FAD-dependent monooxygenase, hydroxylating the δ-amino group of L-Orn, and the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd.

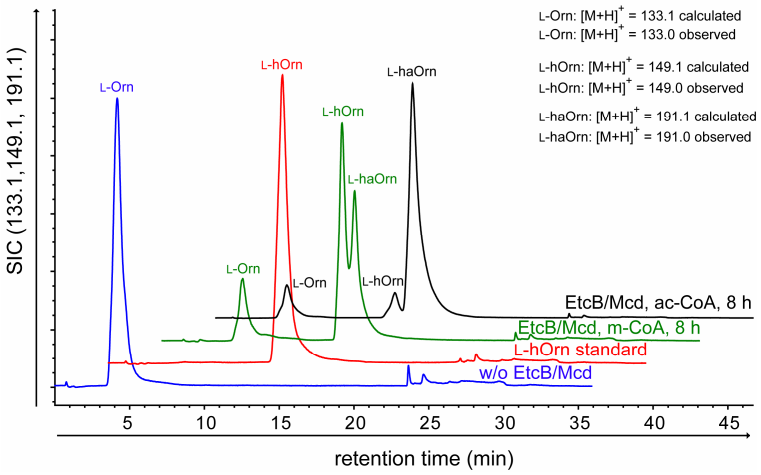



**Figure 5.32:** Consecutive enzymatic synthesis of L-haOrn. The LCMS-traces of the coupled assays are shown as selected ion chromatograms (SIC). Incubation of the substrate L-Orn ($t_R$ = 4.5) with EtcB, Mcd and malonyl-CoA gives rise to the products L-hOrn ($t_R$ = 12.2) and L-haOrn ($t_R$ = 13.3) (green trace). Utilization of acetyl-CoA as acetyl-group donor enhances the generation of L-haOrn, resulting in almost complete substrate turnover (black trace). The control lacking both enzymes is shown in blue and the L-hOrn standard is shown in red. The calculated and observed m/z-ratios are given in the inset.

# 6. Discussion and Outlook

## 6.1 Iterative Assembly of Chromodepsipeptide Derivatives

The exploitation of the macrocyclization potential inherent in TEs dissected from their corresponding nonribosomal peptide synthetases has enabled the generation of novel macrocyclic bioactive compounds, based on the primary sequence of the native substrate, under stringent stereo- and regioselective control.[118] Among the class of nonribosomally synthesized peptides the chromodepsipeptides represent a multitude of structurally and functionally diverse compounds. With the biochemical characterization of TioS PCP-TE a model system for the biocombinatorial synthesis of the quinoline- and quinoxaline-type class of chromodepsipeptides was established. In contrast to linearly operating TEs, TioS PCP-TE acts as an iterative ligation and macrocyclization platform capable of catalyzing macrolactonization and a to date unreported macrothiolactonization. Substrate tolerance of the recombinant enzyme enabled the chemoenzymatic synthesis of chromodepsipeptide analogs, which were evaluated for DNA-bisintercalation activity.

### 6.1.1 TioS PCP-TE-Mediated Dimerization and Macrothiolactonization

Iteratively operating thioesterases catalyze the cyclodimerization of two or more repetitive oligopeptides to afford the readily assembled product. The recently characterized gramicidin S thioesterase GrsB TE confirmed the TE to act as a ligation- and cyclization catalyst following the backward mechanism.[112] Biochemical characterization of TE-domains *in vitro* showed the enzymes to catalyze macrolactonization- and macrolactamization reactions.[53,118,239] Within the study presented herein a set of tetrapeptidyl thioesters was synthesized and incubated with the recombinant cyclase for the investigation of TE-mediated macrothiolactonization (Chapter 5.1.2). The results of this first approach are summarized in Table 6.1.

**Table 6.1:** Overview of the substrates investigated towards macrothiolactonization. The compound names, corresponding peptide sequences, observed products and hydrolysis to cyclization (hy:cy)-ratios are given. The ligation product resembles the *C*-terminally SNAC-activated linear octapeptidylthioester.

| compound | sequence | ligation | cyclization | hy:cy-ratio |
|---|---|---|---|---|
| TL1 | QA-D-$Cys_1$-$Gly_2$-L-$Cys_3$-L-$Cys_4$-SNAC | ✗ | ✗ | / |
| TL2 | QA-D-$Cys_1$-$Gly_2$-L-$Cys_3$-*S*-Me-L-$Cys_4$-SNAC | ✗ | ✗ | / |
| TL3 | QA-D-$Cys_1$-$Gly_2$-L-$Ala_3$-*S*-Me-L-$Cys_4$-SNAC | ✗ | ✓ | 1:7 |
| TL4 | QA-D-$Cys_1$-$Gly_2$-L-$Ala_3$-L-$Met_4$-SNAC | ✓ | ✓ | 1:2 |
| TL5 | QA-D-$Cys_1$-$Gly_2$-L-$Cys_3$-L-$Met_4$-SNAC | ✗ | ✗ | / |
| TL6 | QA-L-$Cys_1$-$Gly_2$-L-$Ala_3$-L-$Met_4$-SNAC | ✗ | ✗ | / |
| TL11 | QX-D-$Cys_1$-$Gly_2$-L-$Ala_3$-*S*-Me-L-$Cys_4$-SNAC | ✗ | ✓ | 1:6 |

At first TioS PCP-TE was tested using a linear tetrapeptide based on the amino acid sequence derived from the specificity prediction of the corresponding A domains. Incubation of TL1 with the thioesterase resulted solely in hydrolytic cleavage of the *C*-terminally SNAC activated thioester. This result led to the conclusion that the steric demand of the *C*-terminal amino acid is crucial for suppression of hydrolysis by shielding the acyl-O-TE oxoester intermediate from the nucleophilic attack of water. Presuming that *S*-methylation of the naturally occurring *S*-Me-L-$Cys_4$ is carried out prior to recognition, activation and incorporation of the building block into the oligopeptide chain, TL2 was synthesized and employed in the macrocyclization assay. Under these conditions hydrolysis was reduced with little substrate remaining after 2 h of incubation, in contrast to total substrate conversion in the case of TL1, confirming the assumption made concerning hydrolysis suppression by steric demand. Dimerization or cyclization was also not observed employing substrate TL2. It was postulated that the sterical demand of the $Cys_3$-residue impedes the cyclodimerization reaction. To evaluate the influence of this cysteine residue on macrothiolactonization, TL3 was employed harbouring an L-$Ala_3$ residue to maintain stereochemical information and concurrently reduce the electrostatic repulsion effects of two neighbouring sulfhydryl groups. Detection of the macrocyclic product indicated a strong influence of this position onto the ligation and cyclization reaction. The reduction of macrocyclization due to the sterical hindrance of two sulfhydryl groups in close proximity has also been observed during the characterization of the echinomycin Ecm7 TE.[240] Linear octapeptidyl-SNAC substrates containing twin thiol groups cyclized with reduced efficiencies, whereas the same octapeptidyl substrates harbouring intramolecular disulfide connectivity showed superior cyclization efficiencies.

In the assembled native thiocoraline the sulfhydryl groups of L-$Cys_3$ form a disulfide crossbridge minimizing conformational freedom to a great extent.[68] It can be assumed that the oxidative formation of the crossbridge is carried out on the PCP-bound linear octapeptidyl-thioester resulting in a prefold facilitating subsequent macrocyclization. This assumption is in compliance with the backward mechanism proposed for iteratively working thioesterases in which the PCP-domain serves as a holding bay for the dimerized product. In the case of echinomycin biosynthesis an oxidoreductase (Ecm17) is found within the biosynthetic operon proposed to be responsible for disulfide formation.[241] This oxidoreductase, although lacking in the gene cluster enabling thiocoraline biosynthesis, could carry out the online modification of the linear PCP-bound octapeptide *in trans*.[136] In contrast to this mechanism, Watanabe and coworkers proposed the putative disulfide forming enzyme Ecm17 to mediate disulfide crossbridging after release of the octapeptidyl macrolactone precursor from the synthetase Ecm7.[241] Ecm17 represents a FAD-dependent pyridine nucleotide-disulfide oxidoreductase requiring $NADP^+$ for

oxidative disulfide formation. This disulfide crossbridged echinomycin precursor is subsequently converted into the corresponding thioacetal linked echinomycin by the SAM-dependent methyltransferase Ecm18 as proven by studies with the recombinant enyme *in vitro* (Figure 6.1).[241] A further example of disulfide bridge forming enzymes mediating the postsynthetic modification of natural products is the FAD-dependent pyridine nucleotide-disulfide oxidoreductase DepH, involved in the formation of the disulfide-linkage found in the antitumor compound FK228.[242] As the cytosolic environment provides reducing conditions, the timing of disulfide crossbridging remains to be elucidated.

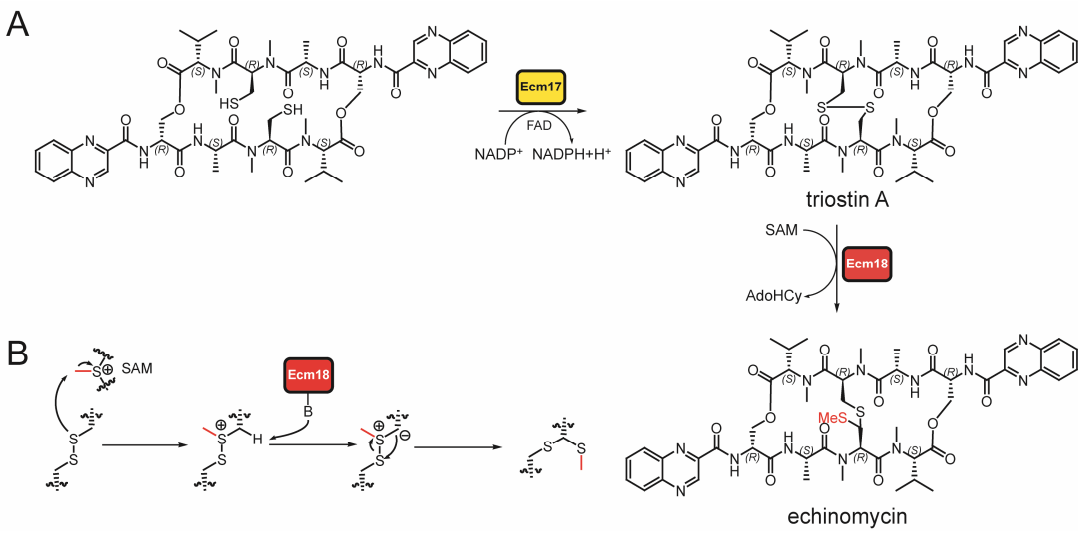



**Figure 6.1:** The postulated model for the generation of the thioacetal crossbridge in echinomycin. A) The octapeptidyl macrolactone is released from the NRPS Ecm7 via TE-mediated cyclodimerization of tetrapeptidyl substrates. Disulfide crossbridging is governed by Ecm17 affording triostin A. Ecm18 catalyzed thioacetal formation gives rise to echinomycin. B) The mechanism for Ecm18 mediated methylation of the disulfide. In a first step, Ecm18 catalyzes the nucleophilic attack of sulphur onto the methyl group of SAM. Subsequently, base assisted deprotonation and intramolecular reorganization results in the thioacetal group.

To further prove that the steric demand of the *C*-terminus is a key position in thiocoraline macrothiolactonization, *S*-Me-L-$Cys_4$ was substituted with L-$Met_4$, resulting in an improved hydrolysis to cyclization-ratio of 1:2 for TL4 (TL3 12:1, 25°C), reinforcing former presumptions. Intriguingly, the substitution also led to the buildup of a linear ligation product which can be directly assigned to the backward mechanism.

Following this mechanism, the TE-bound tetrapeptide undergoes a nucleophilic attack by the external tetrapeptidyl-SNAC mimicking the PCP-bound tetrapeptide. The *C*-terminal steric demand then inhibits

subsequent macrocylization and leads to the accumulation of the octapeptidyl-SNAC, resembling the naturally occurring PCP-bound octadepsipeptide. These observations directly correlate with the results obtained with GrsB PCP-TE and confirm TioS PCP-TE to follow the backward mechanism logic (Figure 6.2).[112] This backward mechanism has now been confirmed for three individual thioesterase domains catalyzing macrolactonization (echinomycin), macrolactamization (gramicidin S) and macrothiolactonization *in vitro*.[20,112,240]

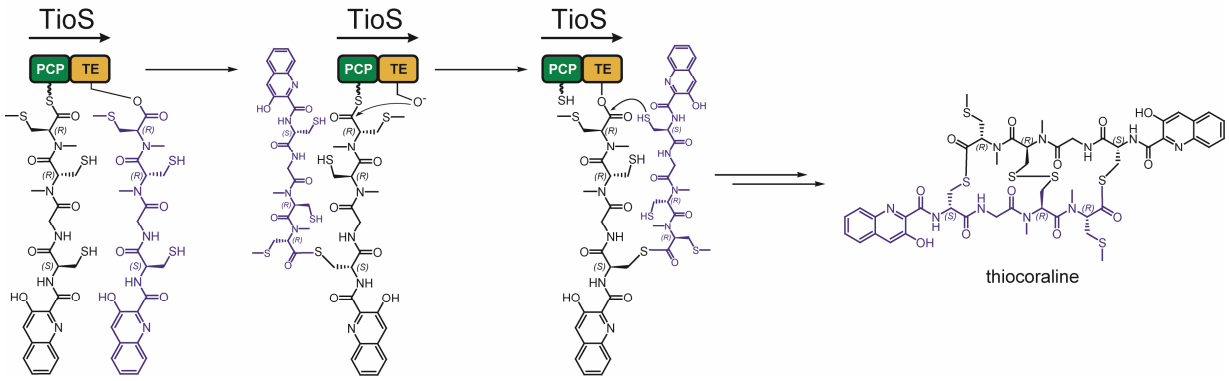



**Figure 6.2:** Mechanistic model of the TioS PCP-TE catalyzed cyclodimerization reaction during thiocoraline biosynthesis *in vivo*. Ligation and subsequent cyclization follows the backward mechanism logic. The PCP-bound tetrapeptide conducts a nucleophilic attack onto the acyl-O-TE oxoester of the TE-bound tetrapeptide (accentuated in blue) resulting in a PCP-bound linear octathiodepsipeptide. This PCP-bound intermediate is then transferred onto the active-site serine of the adjacent TE-domain. Subsequent nucleophilic attack of a free thiol group gives rise to the cyclic thiodepsipeptide which is modified by means of disulfide crossbridging.

All known chromodepsipeptides share a D-configured *N*-terminal amino acid which is harbouring the nucleophilic side-chain that mediates cyclization.[124] Substitution of this position with L-$Cys_1$ (TL6) abolished ligation and subsequent cyclization resulting in complete hydrolysis. This observation indicates that only D-configured amino acids enable the specific angle, following Bürgi-Dunitz trajectory, required for the nucleophilic attack onto the acyl-O-TE oxoester intermediate.[243] Furthermore, the correct positioning of the substrate within the catalytic pocket of the thioesterase might be influenced. This result is also in full agreement with the biochemical characterization of several NRPS-derived TE-domains *in vitro*. During the characterization of the surfactin or the CDA cyclase.the stereoconfiguration of the *C*- or the *N*-terminal amino acid was found to be essential for macrocyclization.[53,180]

In contrast to the native thiocoraline tetrapeptide, all utilized peptidyl substrates lacked *N*-methylation of residues 3 and 4 and the hydroxylation of the chromophore moiety. Alkylations of amide-bonds shift

the equilibrium between *cis*- and *trans*-conformation towards *cis*, influencing the overall conformation of the peptide.[244] Furthermore, rotational barriers are elevated with the peptide being rigidified in a specific conformation.[245] This missing structural restriction may also be responsible for the observed hydrolysis rates or the lack of substrate conversion.

### 6.1.2 Biocombinatorial Potential of TioS PCP-TE

To further investigate the biocombinatorial potential of TioS PCP-TE, a set of D-$Ser_1$ substituted tetrapeptidyl-SNACs (TL7-TL10) was analyzed (Chapter 5.1.3). The results of these experiments are summarized in Table 6.2.

**Table 6.2:** Overview of the substrates investigated towards macrolactonization. The compound names, corresponding peptide sequences, observed products and hydrolysis to cyclization (hy:cy)-ratios are given. The ligation product resembles the *C*-terminally SNAC-activated linear octapeptidylthioester.

| compound | sequence | ligation | cyclization | hy:cy-ratio |
|---|---|---|---|---|
| TL7 | QA-D-$Ser_1$-$Gly_2$-L-$Cys_3$-L-$Cys_4$-SNAC | x | ✓ | 5:1 |
| TL8 | QA-D-$Ser_1$-$Gly_2$-L-$Cys_3$-*S*-Me-L-$Cys_4$-SNAC | x | ✓ | 1:4 |
| TL9 | QA-D-$Ser_1$-$Gly_2$-L-$Ala_3$-*S*-Me-L-$Cys_4$-SNAC | x | ✓ | 4:1 |
| TL10 | QA-D-$Ser_1$-$Gly_2$-L-$Ala_3$-L-$Met_4$-SNAC | ✓ | ✓ | 8:1 |

In contrast to TL1, the serine-substituted TL7 was cyclized leading to the conclusion, that L-$Cys_3$ is only influencing macrothiolactonization. In this case the electrostatic repulsion effects only occur when L-$Cys_3$ of one tetrapeptide and D-$Cys_1$ of the other peptide chain are in close proximity. With the sterically less demanding D-$Ser_1$, macrolactonization is feasible even in presence of L-$Cys_3$. When incubating TioS PCP-TE with TL8 the formation of a four residue macrolactone (Cy8/4) was detected (Figure S1). This macrolactonization of a single tetrapeptidyl-SNAC was solely observed with TL8. In contrast to TL7 the *C*-terminal steric demand is leading to a more stable TE-bound intermediate allowing an intramolecular attack of D-$Ser_1$ onto the acyl-O-TE oxoester intermediate prior to hydrolytic cleavage. Alteration of the *C*-terminal chromophore from quinoline to quinoxaline does not influence cyclization yields to a great extent and allows the generation of quinoxaline-type chromodepsipeptides. In summary, TioS PCP-TE is the first dissected cyclase catalyzing both macrothiolactonization and macrolactonization.

### 6.1.3 Concepts for the Improvement of Cyclization Yields

Enzymatic peptide cyclization often displays low efficiency due to the occurrence of hydrolysis of the acyl-O-TE oxoester intermediate. Previous work on the excised TE-domains from the tyrocidine and pristinamycin synthetases revealed hydrolysis to cyclization-ratios of 1:1 and 1:3 for natural substrate analogs.[246-247] Macrocyclization assays described herein also revealed a high degree of hydrolysis typical for some isolated TE-domains. In order to improve cyclization yields the temperature dependence of either macrothiolactonization or macrolactonization was evaluated. TE-mediated macrothiolactonization represents an energetically less favored reaction due to the fact that a thermodynamically stable oxoester is converted to a high energy thioester.

Increasing the temperature also increased the formation of the endergonically generated macrothiolactone in the case of TL3 and TL11 (Chapter 5.1.4). A reduction of the temperature also resulted in the increase of cyclization yields. It was speculated that low temperatures induce a more compact conformation of the enzyme. Under these conditions premature hydrolysis is reduced increasing the stability of the acyl-O-TE oxoester intermediate capable of reacting with further molecules to give rise to the macrocyle. In contrast, the thermodynamically indifferent macrolactonization is favored at low temperatures utilizing TL9. Analogous results were obtained with substrate TL8. In all examined cases total substrate conversion is decelerated at lower temperatures reflecting minimized reaction velocities. Kinetic investigation of TL3 turnover resulted in a $k_{cat}$ of 5.26 ± 0.64 $min^{-1}$ being in range of the corresponding substrate turnover of the linear pentapeptidyl-thiophenol of GrsB PCP-TE ($k_{cat}$ = 2.4 $min^{-1}$).[112] Substrate turnover of TL8 is given by a $k_{cat}$ of 8.92 ± 1.2 $min^{-1}$. The higher $k_{cat}$-value for TL8 is the result of an increased flux towards hydrolysis compared to TL3 and does not resemble an improved cyclization efficiency. Higher catalytic efficiencies can be expected when the linear octapeptide is used due to the fact that the ligation reaction is the rate-determining step as described for GrsB PCP-TE.[112] Additional studies will have to be conducted to prove the generality of the approach described herein. Methods for the reduction of the hydrolysis ratio such as the addition of non-ionic detergents to the reaction mixture or the utilization of ionic liquids or organic solvents have also been reported.[247-248]

An alternate approach to increase the yield of macrocyclic products was established with the recombinant TE-domain dissected from Ecm7, a synthetase involved in the assembly of the macrolactone echinomycin.[240] The cyclase was incubated with several tetrapeptidyl- and octapeptidyl thioesters resembling the native sequence of echinomycin in the absence or presence of DNA. It was found that hydrolysis was reduced to a great extent in the presence of DNA. Furthermore it was shown that the addition of DNA improved the yields of macrolactones with a complete reversal of the hydrolysis

to cyclization-ratio for several substrates especially for TANDEM (2:1 →1:18) (Figure 6.3).[240] Product inhibition and TE-mediated hydrolysis of the macrocycles is circumvented by the capture of the DNA-bisintercalating compound with double-helical DNA containing bisintercalation sites. This chemoenzymatic generation of chromodepsipeptides in the presence of DNA represents an effective method for the screening of DNA-binding small molecules by coupling the production of the compound with a preliminary screening of DNA-binding ability and can be applied for the generation of thiocoraline analogues.

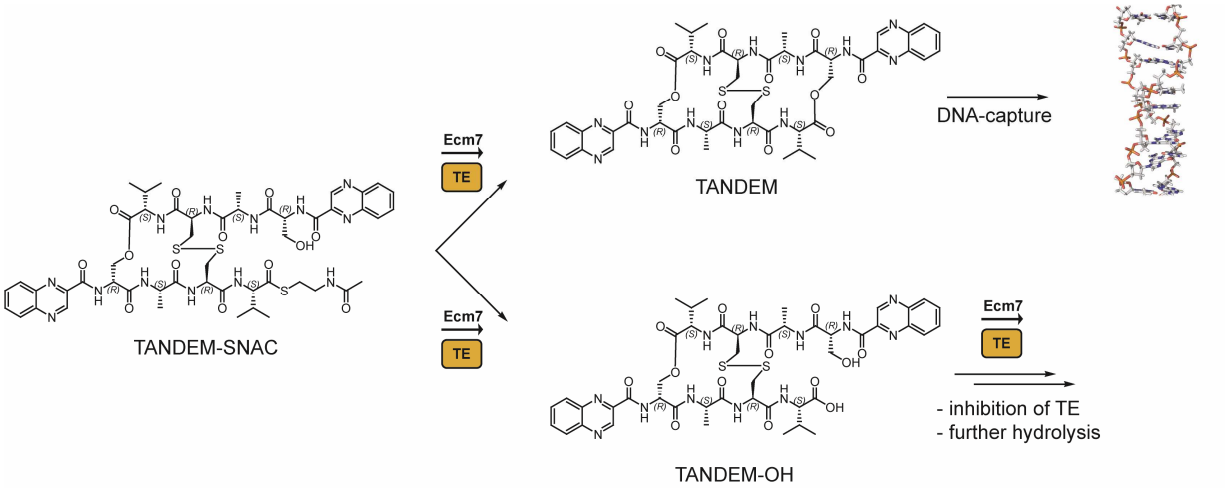



**Figure 6.3:** Schematic overview of the effect of DNA on the Ecm7 TE-mediated macrocyclization of the substrate TANDEM-SNAC. Absence of DNA gives rise to the hydrolysis product TANDEM-OH, which is furthermore hydrolytically cleaved and inhibits Ecm7 TE. In contrast, presence of DNA leads to a reversal of the hydrolysis to cyclization-ratio (2:1 → 1:18) and prevents further hydrolytic cleavage via DNA-capture processes giving rise to the desired product TANDEM.

The mechanisms of how iteratively operating thioesterases can control the number of repetitive ligation steps is yet unknown. Throughout all cyclization reactions the ring sizes of the resulting macrocycles were limited to four residue rings (Cy8/4) or to eight residue rings. In contrast, GrsB PCP-TE is capable of trimerizing pentapeptidyl-SNAC substrates to form macrocycles composed of up to 15 residues.[112] It could not be determined if the capacity of the active-site is the size determining factor. Preorganization and structure of the substrate is also considered to determine whether a ligation or a cyclization step is carried out. Macrocyclization studies employing the recombinant TycC TE, derived from the tyrocidine biosynthetic pathway, proved preorganization of the substrates to be crucial for cyclization efficiency.[249] NMR-structures of the decapeptides tyrocidine and gramicidin S proved both compounds to exhibit a β-sheet like antiparallel fold (Figure 6.4).[249-251] It was proposed that the prefold of the linear substrates

enhances the catalytic efficiency of the TE-mediated macrocyclization due to close spatial coordination of amino acids directly participating in the cyclization reaction.[59] Prefold-induced spontaneous macrocyclization was also observed during the autocatalytic macroimination of nostocyclopeptide. Unfortunately, the prefold of the linear thiocoraline octapeptide has not yet been investigated. In addition, it is assumed that 12-residue rings could exceed the maximum capacity of the catalytic pocket.

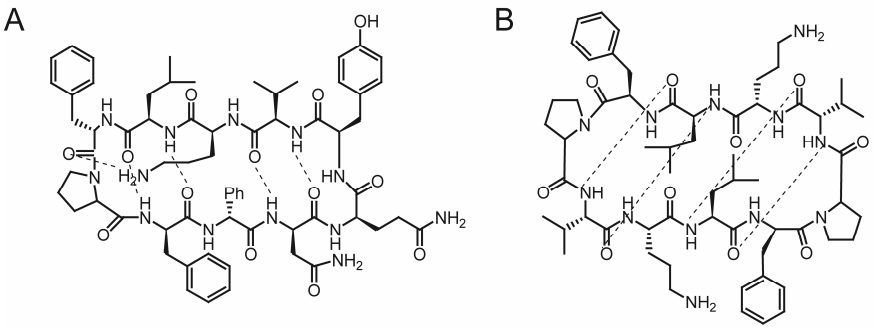



**Figure 6.4:** A network of hydrogen-bonds constitutes the antiparallel fold of the cyclic macrolactams tyrocidine and gramicidin S. A) Five hydrogen bonds give rise to the β-sheet structure of tyrocidine as proposed from the cyclization efficiency of backbone-substituted tyrocidine analogues. B) β-sheet fold of gramicidin S is induced by four hydrogen bonds and was determined via NMR.

In conclusion, the excised thioesterase of thiocoraline is a versatile catalyst for the *in vitro*-generation of chromodepsipeptide analogs. TioS T-TE is the first cyclase characterized capable of catalyzing macrothiolactonization. Additionally, macrolactonization is feasible due to relaxed substrate specificity towards the cyclizing nucleophile. Cyclization yields can be improved by temperature shifts or by utilizing methods already reported.[240,247-248] Substrate tolerance towards the chromophore moiety also allows the chemoenzymatic synthesis of quinoxaline substituted analogs mimicking the class of triostins and echinomycins. The approach presented herein provides new opportunities to develop novel compounds related to thiocoraline and similar oligopeptides with a potentially improved spectrum of pharmacologicalproperties and higher *in vivo*-stability. Future studies will aim at the generation of amide-linked thiocoraline derivatives by substitution of the D-$Cys_1$ residue with D-$Dap_1$ (D-diaminopropionate).

### 6.1.4 DNA-Bisintercalative Activity of Thiocoraline Analogs

Thiocoraline attains the antiproliferative activity from its capability to bisintercalate into duplex DNA, leading to the inhibition of DNA-polymerase α.[141] Although no three-dimensional structure of thiocoraline intercalated into DNA is available, it is postulated that thiocoraline binds to DNA in a mode similar to echinomycin.[134,252] The three-dimensional structure of echinomycin intercalated into duplex DNA shows the molecule to adopt the same U-shaped conformation upon intercalation into DNA as postulated for thiocoraline based on molecular dynamics simulations (Figure 6.5).[252] In this cocrystal structure, the quinoxaline-moieties (QX) enable π-π interactions with the aromatic nucleobases, distorting the helical structure of the DNA.

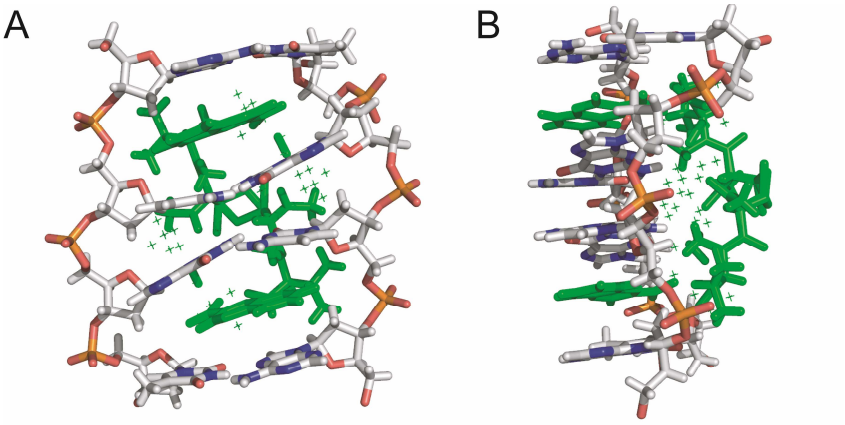



**Figure 6.5:** A) The three-dimensional structure of echinomycin intercalated into duplex DNA (5'-ACGTACGT-3') (PDB code: 2ADW). Echinomycin is presented in green. The chromophore moieties ensure π-π-stacking effects inducing a U-shaped conformation of the bisintercalator. Echinomycin is shown to bind to the minor groove of the DNA encompassing a central CpG motif. An analogous bisintercalation mode has been postulated for thiocoraline based on molecular dynamics simulations. B) The same structure rotated by 90° along the y-axis.

To investigate the potential bioactivity of the generated macrocycles, several were isolated and employed in a DNA-bisintercalation activity assay (Chapter 5.1.5) (Figure 6.6). Thiocoraline stabilized duplex DNA in a range similar to previously described results, whereas bisintercalation of the analogs could not be detected.[134] The generated thiocoraline analogs displayed a variety of modifications of the peptidic backbone compared to the native bisintercalator. The glycine residue was conserved throughout the analogs as it is postulated that hydrogen-bonds are formed between the NH- and carbonyl groups of glycine and the N3- and exocyclic amino group of guanine.[134]

Substitution of the naturally occurring chromophore moiety 3HQA with QA or QX is not likely to affect the bisintercalation properties. QX is found in the well characterized DNA-bisintercalators echinomycin and triostin A, nevertheless Cy11 did not show any activity (Figure 6.6). Furthermore QA-substituted chromodepsipeptides, belonging to the recently synthesized FAJANU peptide family, also showed bioactivity against several tumor cell lines *in vivo*.[253] FAJANU 7, a QA-capped eight residue macrolactam displayed highest bioactivity even exceeding 3HQA or QX harbouring compounds. The lack of *N*-methylation of L-$Cys_{3/4}$ is presumably responsible for the absence of DNA-bisintercalation activity. *N*-methylation induces conformational changes and elevates rotational barriers.[245] This rigidification of molecular dynamics gives rise to a preferential prefold of the oligopeptide. The substitution of *N*-Me-Gly residues with Gly in the case of FAJANU chromodepsipeptides led to the decrease of bioactivity by one order. In addition, the intramolecular disulfide connectivity is also likely to induce a preorganization of the molecule facilitating DNA-bisintercalation. This structural rigidification was only present in the macrocycle Cy8SS but did not affect the stabilization of DNA. The choice of the oligonucleotide has also to be considered as the oligonucleotide AS displayed the highest degree of stabilization upon incubation with thiocoraline, but might represent the inappropriate template for the investigation of thiocoraline analogue DNA-bisintercalation. Obviously, additional extensive studies will be necessary to gain further insights into the molecular mechanism of thiocoraline bioactivity and its MOA to rationally generate chromodepsipeptide analogs.

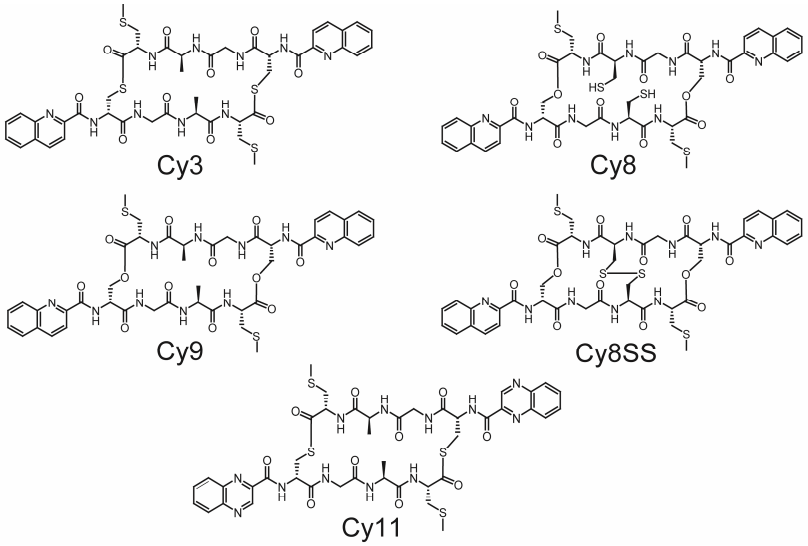



**Figure 6.6:** The structures of the chemoenzymatically generated macrocycles investigated for DNA-bisintercalation activity.

## 6.2 Erythrochelin – a Hydroxamate-Type Siderophore Discovered via Genome Mining

The genome of the erythromycin-producing strain *Saccharopolyspora erythraea* contains 25 secondary metabolite gene-clusters mostly considered to be orphan including two responsible for siderophore assembly. Within the study presented herein, the isolation and structural elucidation of the hydroxamate-type tetrapeptide siderophore erythrochelin, the first NRPS-derived natural product of *S. erythraea*, was reported. In an attempt to substitute the traditional activity assay-guided isolation of novel secondary metabolites, a dedicated radio-LCMS-methodology to identify NRPs of cryptic gene clusters in the industrially relevant strain was established. This methodology was based on transcriptome data and adenylation domain specificity prediction and resulted in the detection of a radiolabeled ornithine-inheriting hydroxamate-type siderophore. The improvement of siderophore production enabled overall structural elucidation via NMR- and $MS^n$-analysis and hydrolysate-derivatization for the determination of amino acid configuration. The sequence of the tetrapeptide siderophore erythrochelin was determined to be α-*N*-acetyl-δ-*N*-acetyl-δ-*N*-hydroxy-D-ornithine-D-serine-cyclo(δ-*N*-hydroxy-L-ornithine-δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine). Bioinformatic analyis of the S*. erythraea* genome revealed the enzyme responsible for δ-*N*-acetylation of L-hOrn to be encoded in a distant NRPS gene cluster. Biochemical characterization of this bifunctional malonyl-CoA decarboxylase/*N*-acetyltransferase together with two FAD-dependent monooxygenases established a biosynthetic route for the generation of the L-haOrn building block found in erythrochelin. The results derived from structural and functional characterization of erythrochelin, together with the biochemical characterization of the tailoring enzymes EtcB, Sace_1309 and Mcd, enabled the proposal of a biosynthetic pathway for erythrochelin biosynthesis. In this model the tetrapeptide is assembled by the tetramodular NRPS EtcD, involving unusual initiation- and cyclorelease-mechanisms.

### 6.2.1 Natural Product Discovery via Radio-LCMS-Guided Genome Mining

Advances in DNA-sequencing technologies by the development of high-throughput pyrosequencing (454, SOLEXA) gave rise to over 1000 sequenced microbial genomes (Genomes Online Database). Automated annotation of the genomes unveiled a large number of orphan (cryptic) gene clusters related to the biosynthesis of secondary metabolites representing a rich resource for the discovery of bioactive natural products. By applying systematic state-of-the-art bioinformatic analysis of these biosynthetic clusters structural elements or even the activity of the cryptic product can be predicted. The subsequent isolation of these compounds solely based on genomic sequence data was baptized genome mining.[254] Several strategies were developed to assist the genome mining process being either targeted (heterologous

expression of the gene cluster, gene knockout/comparative metabolic profiling) or general (genomisotopic approach).[39] The radio-LCMS-guided genome mining method established within this study represents a general and rational strategy for the discovery of NRPS-derived natural products. This concept relies on the bioinformatic analysis of the targeted gene clusters for the prediction of the substrates incorporated into the natural product. These building blocks are then fed to cultures of the target strain in an isotopically labeled form and incorporated into the natural product, which is subsequently detected with a radioactivity flow-through detector coupled to a HRMS instrument (Figure 6.7).

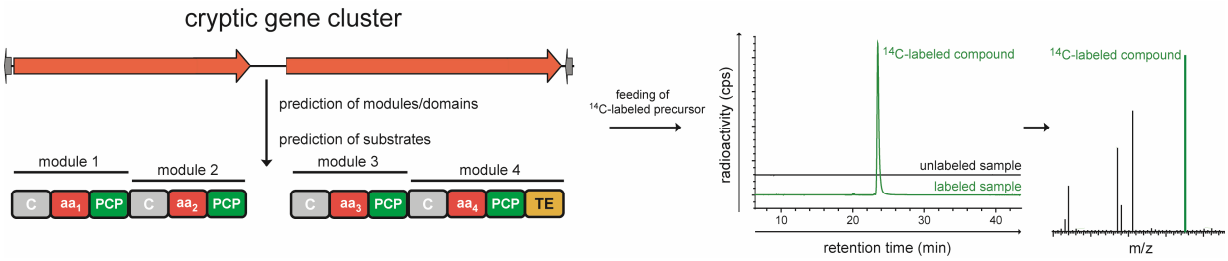



**Figure 6.7:** The concept of natural product discovery via radio-LCMS-guided genome mining. Bioinformatic analysis of the cryptic genes leads to the prediction of putative substrates. Feeding of the precursor as $^{14}$C-labeled molecule gives rise to the isotopically labeled compound which is identified with a radioactivity flow-through detector coupled to a HRMS-instrument.

Analysis of the *nrps3* and the *etc* gene cluster identified L-Orn to be incorporated in both putative siderophores. In addition, transcriptome analysis of *Saccharopolyspora erythraea* NRRL 23338 cultures grown in SCM-medium confirmed upregulation of both clusters in growth phases B/C.[225] Initial detection of a CAS-positive compound was performed by cultivation of *S. erythraea* in SCM-medium, confirming DNA-microarray gene expression profiles obtained for *S. erythraea*. LCMS-analysis of this CAS-positive fraction revealed several compounds coeluting, thus impeding the direct MS-based identification of the target compound (Figure 6.8 A). In order to circumvent this activity-based genome mining approach, the radio-LCMS-guided concept was carried out. Feeding of the nonproteinogenic amino acid $^{14}$C-L-Orn prior to expression of the *etc* and *nrps3* gene clusters gave rise to a radiolabeled compound, which could be clearly identified on an analytical scale and displayed the same m/z-ratio as one of the compounds in the CAS-positive fraction (Figure 6.8 A). The sensitivity of radioactivity detection and sophisticated analytical separation proved to be advantageous in this approach. The iron-chelative properties of the radiolabeled

compound were confirmed by CAS-assay-guided fractionation of medium-scale fermentation extractions and the detection of a radiolabeled *ferri*-siderophore complex (Figure 6.8 B).

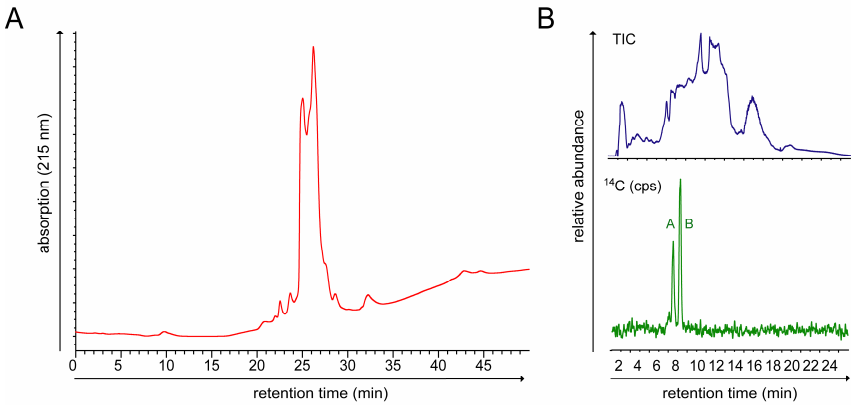



**Figure 6.8:** The sensitivity of radio-LCMS guided natural product discovery. A) LCMS-analysis of the fraction identified via CAS-activity assay. Fraction complexicity as detected by UV/Vis-analysis impedes the identification and isolation of the target compound. B) Radio-LCMS-guided identification of the target compound. The TIC (blue) exemplifies the complexicity of the sample. The radioactivity flow-through detector (green) reveals two radiolabeled compounds in the complex sample and enables a direct correlation between the radioactivity signal and high resolution mass signal. Compound A is *ferri*-erythrochelin and compound B represents erythrochelin.

Cultivation of *S. erythraea* under iron-depleted conditions induced the production of erythrochelin compared to iron-rich media cultivations. This observation has been described for the isolation of several siderophores including fuscachelin and coelichelin and involves iron-sensing processes by the microorganisms.[37-38]

Interestingly, the amount of $^{14}$C-L-Orn incorporation was increased from 2% to 4% (based on the total amount of radioactivity fed) when switching to minimal media. It is likely that the decelerated growth in iron-depleted minimal media combined with an increase in siderophore production leads to the increased incorporation of $^{14}$C-L-Orn into the main secondary metabolite erythrochelin. The utilization of radiolabeled proteinogenic amino acids, which can be channelled to ribosomal synthesis of peptides or to the primary metabolism, remains to be elucidated. Throughout the experiments, the second catecholate/hydroxamate-type siderophore assembled by the NRPSs encoded in *nrps3* could not be detected applying the aforementioned approach (Figure S8). In addition, the extraction of culture supernatants of *S. erythraea*, cell pellets and lysed cells with a variety of organic solvents did not lead to the identification of a second CAS-reactive compound (data not shown). It is therefore assumed that

either the extraction conditions were inadequate for the isolation of the natural product, or that the gene cluster is silent under the conditions employed**.** It is also postulated that *S. erythraea* might assemble the second compound to serve as a backup siderophore upon iron starvation.[238]

In conclusion, the described approach, solely based on A-domain specificity prediction and available transcriptome data, can be applied for the initial detection and isolation of NRPs, enabling the scale-down of NRP discovery from preparative to analytical scale. Radio-LCMS-guided genome mining can be utilized to substitute the detection and isolation of NRPs based on their biological activity, which is often challenging to predict. The main advantage of this approach in comparison with the established strategies is the generality of this rapid method as several biosynthetic gene clusters can be targeted in parallel when the isotope label is predicted to be incorporated into different cryptic products. A limitation of this approach is the selection of the appropriate precursor as some are toxic to specific organisms or can be incorporated into other secondary metabolites.[255] In addition, specificity prediction has to be accurate for the label to be incorporated into the product.[196] Degradation of the amino acid can also result in the dispersal of the label to other compounds. Identifying the right cultivation conditions leading to gene cluster expression is also crucial for this strategy to result in the identification of the compound as it does not uncouple the biosynthetic enzymes from native regulatory mechanisms compared to the heterologous gene cluster expression concept.[196,241]

### 6.2.2 Physicochemical Properties of Erythrochelin

The overall structure of erythrochelin was determined by NMR- and MS-analysis as well as hydrolysate-derivatization for determination of amino acid configuration. The peptide sequence is composed of D-ac-haOrn$_1$-D-Ser$_2$-cyclo(L-hOrn$_3$-L-haOrn$_4$). Erythrochelin represents a hydroxamate-type tetrapeptide siderophore containing three ornithine residues of which two are δ-*N*-acetylated and δ-*N*-hydroxylated. In addition, the *N*-terminal α-amino group of haOrn$_1$ is capped with an acetyl moiety. A local symmetry in erythrochelin is attained by a DKP-structure consisting of two cyclodimerized L-Orn residues. The mode of Fe(III)-chelation by erythrochelin remains to be elucidated, but an iron binding mode analogous to gallium-binding by coelichelin is postulated and has recently been described (Figure 6.9).[38,238] The NMR spectroscopic analysis of the Ga(III)-erythrochelin complex showed an upfield shift for the serine carbonyl group in addition to upfield shifts of the carbonyl groups of D-ac-haOrn$_1$ and L-hOrn$_3$. Analogous correlations have been reported for the Ga(III)-coelichelin complex.[38] MS-analysis of *ferri*-erythrochelin reveals an abolished skimmer fragmentation compared to erythrochelin, being

indicative of an induced rigidification of the siderophore upon iron-binding. Erythrochelin shows an absorption spectrum typical of *ferri*-hydroxamate siderophores with $\lambda_{max}$ = 440 nm.

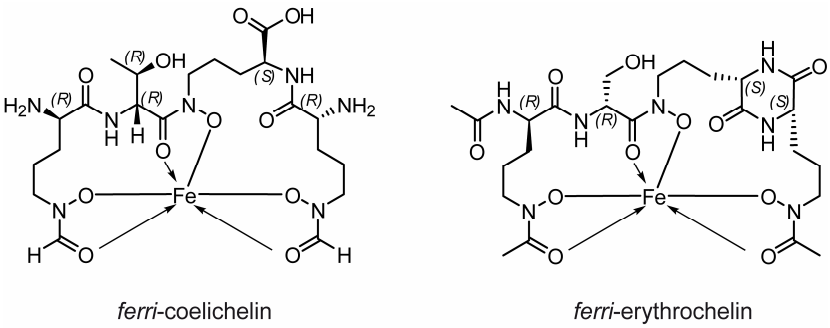



**Figure 6.9:** Proposed Fe(III)-binding modes of coelichelin compared to erythrochelin. The proposed model is supported by MS-data of *ferri*-erythrochelin and NMR analysis of Ga(III)-erythrochelin. Coordination of iron involves three carbonyl- and three hydroxyl-groups.

The structure of erythrochelin resembles that of DKP-containing fungal siderophores exemplified by coprogen (*Pezizomycota* species) or rhodotorulic acid (*Rhodotorula basiodomycetes*), which solely consists of two cyclodimerized *N*-acetyl-*N*-hydroxyornithine residues (Figure 6.10).[256-257] Erythrochelin shares a high degree of structural similarity to the angiotensin-converting enzyme (ACE)-inhibitor foroxymithine, isolated from cultures of *Streptomyces nitrosporeus* (Figure 6.10).[229-230] In contrast to erythrochelin, the δ-amino groups of ac-$hOrn_1$ and $hOrn_4$ are formylated suggesting a formyltransferase to be involved in the biosynthesis, similar to coelichelin assembly.[38] In an attempt to chemically obtain foroxymithine, a total synthesis was established by Miller and coworkers, which resulted in a compound exhibiting the same NMR-spectroscopic properties as the isolated natural product. All residues within the peptide chain showed L-configuration.[258] This stereochemistry differs from erythrochelin, in which two residues show a D-configured stereocenter thus suggesting a similar NRPS-based assembly of foroxymithine by a synthetase lacking all E-domains. The lack of sequence information of the *S. nitrosporeus* genome impeded the identification of a biosynthetic machinery governing foroxymithine assembly. Based on the structural similarity to the ACE-inhibitor foroxymithine, ACE-inhibition assays were carried out employing both erythrochelin and *ferri*-erythrochelin. In contrast to foroxymithine, which has been reported as a hypertension reducing agent with an $IC_{50}$-value of 7 µg/mL, erythrochelin does not inhibit the proteolytic cleavage of the artificial substrate HHL.[229] It can only be speculated if the substitution of the δ-hydroxamino acetyl-groups with formyl-groups affects ACE inhibitory activity or if the altered stereochemistry results in abolished *in vitro* activity since no MOA studies of foroxymithine are available.

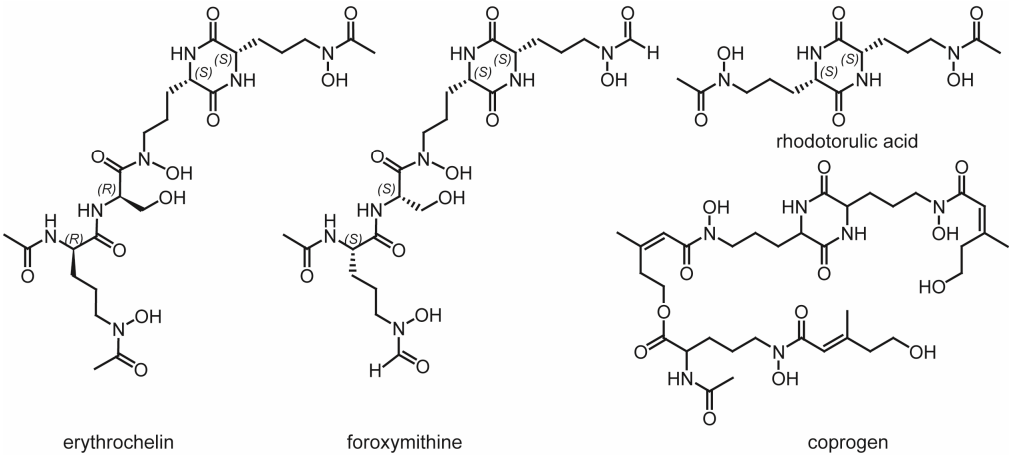



**Figure 6.10:** Chemical structures of the DKP-containing hydroxamate-type siderophores erythrochelin, foroxymithine, rhodotorulic acid and coprogen. The siderophores contain the characteristic diketopiperazine-moiety and tailored ornithine residues involved in the coordination of ferric iron.

## 6.3 Biosynthesis of the Modified Ornithine Residues in Erythrochelin

### 6.3.1 Characterization of Ornithine δ-*N*-Hydroxylation

#### 6.3.1.1 Bioinformatic Analysis of the FAD-Dependent Monooxygenases EtcB and Sace_1309

Biosynthetic analysis of the *etc* and the *nrps1* gene cluster revealed two FAD-dependent monooxygenases, EtcB and Sace_1309, putatively involved in the generation of δ-*N*-hydroxy-L-ornithine required for erythrochelin assembly. Multiple sequence alignments of the two enzymes with the characterized FAD- and NADPH-dependent monooxygenases CchB, IucD, PvdA and VbsO confirmed EtcB and Sace_1309 to share a high degree of sequential identity to this class of enzymes (Figure 6.11).[98-99,169,259-260] Three highly conserved motifs could be identified in the sequence of the monooxygenases of which two are considered to be involved in nucleotide binding. Binding of the cofactor FAD is enabled by the *N*-terminal binding site (GxGxxG), whereas the second binding site (GxGxxG) is responsible for binding of NADPH.[261] Interestingly, the last Gly residue of the FAD binding site is substituted with Pro in FAD-dependent monooxygenases associated with siderophore biosynthesis.[262]

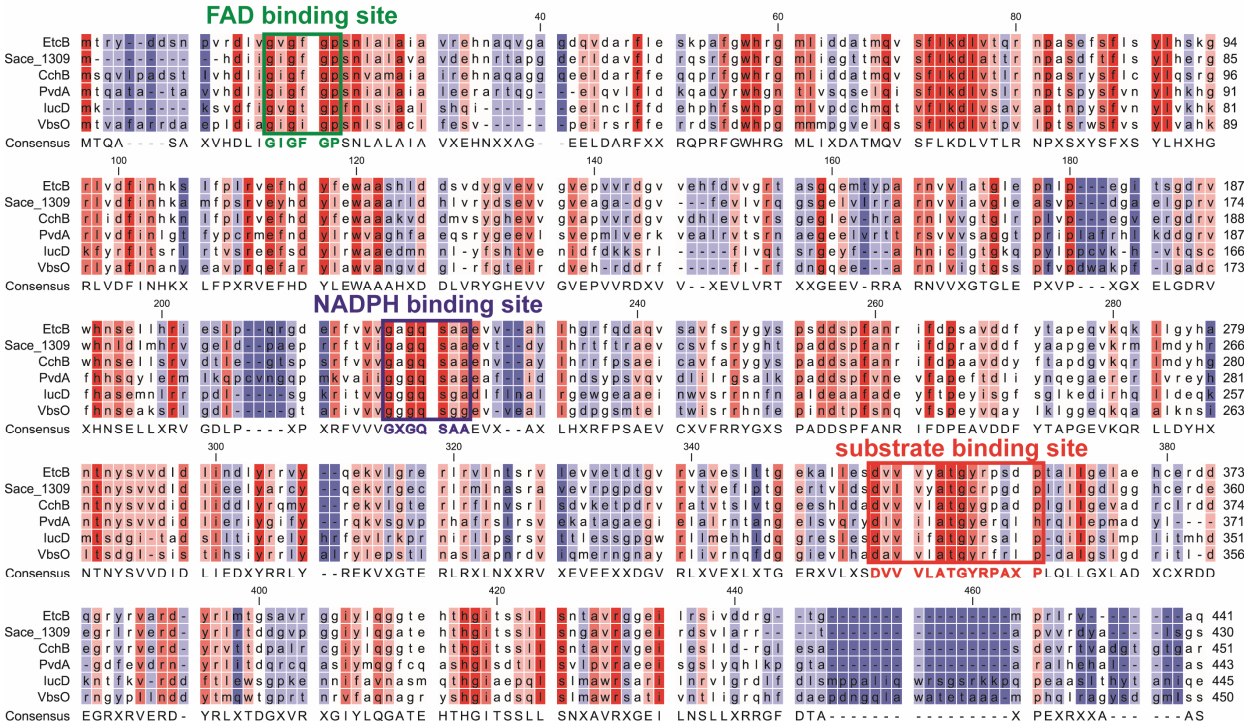



**Figure 6.11:** Multiple sequence alignments of FAD-dependent monooxygenases. Alignments were performed under application of a ClustalW algorithm. The highly conserved *N*-terminal nucleotide binding site responsible for FAD binding is accentuated in green, whereas the motif involved in NADPH coordination is given in blue. The *C*-terminally located substrate binding motif constituting a hydrophobic pocket is accentuated in red. The degree of conservation is indicated by color: Red indicates complete and blue no agreement.

The third conserved motif (DxxxL/FATGYxxxxP), located in the *C*-terminal region of the monooxygenases, is assumed to be involved in substrate binding.[262] The central amino acids L/FATGY constitute a hydrophobic pocket in which the substrate can be coordinated. The conserved aspartic acid residue of this motif is proposed to catalyze the deprotonation of the side-chain ammonium group of L-Lys or L-Orn, enhancing the nucleophilicity and reactivity of this position. EtcB displays 65% sequential identity and 78% sequential similarity to CchB, whereas Sace_1309 shares 58% sequential identity and 70% similarity with CchB.

Spectrophotometric analysis of purified EtcB or Sace_1309 confirmed both enzymes to be partially loaded with the cofactor FAD, which has been reported for several FAD-dependent enzymes. This cofactor is noncovalently associated with the protein in the case of external FAD-dependent monooxygenases. It was shown that recombinant EtcB or Sace_1309 catalyze δ-*N*-hydroxylation of L-Orn in the presence of the cofactor FAD, the reducing cosubstrate NADPH and molecular oxygen, whereas NADH is not accepted as cosubstrate (Figure 6.12). This specificity towards the cofactor and reducing

cosubstrate has also been observed for CchB and PvdA, as both enzymes solely exhibit activity in the presence of FAD and NADPH. In contrast, IucD mediates hydroxylation of L-Lys in the presence of FAD and NADPH or NADH as the reducing cosubstrates.

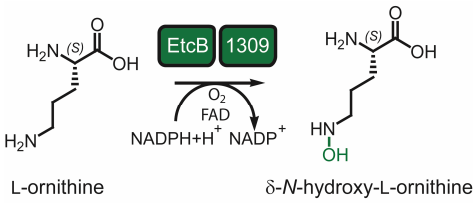



**Figure 6.12:** The FAD-dependent monooxygenases EtcB and Sace_1309 catalyze the hydroxylation of the δ-amino group of L-Orn in the presence of FAD, NADPH and molecular oxygen to afford δ-*N*-hydroxy-L-ornithine.

### 6.3.1.2 Substrate Specificity and Kinetic Parameters of the Monooxygenases

The substrate specificities of EtcB and Sace_1309 were evaluated by incubating the recombinant enzymes with a set of alternative substrates representing the different classes of amino acids. Neither the D-Orn, L-Lys nor α-*N*-acetyl-L-ornithine, representing a possible precursor of the α-*N*-acetyl-δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine$_1$ building block, were accepted. It was shown that the enzymes exclusively catalyze the δ-*N*-hydroxylation of L-Orn. Similar results were obtained during the characterization of FAD-dependent monooxygenases involved in siderophore biosynthesis. IucD specifically hydroxylates L-Lys during aerobactin biosynthesis, whereas hydroxylation activity of CchB or PvdA is exclusively limited to L-Orn.[98-99,169,259-260] Kinetic parameters for the EtcB-mediated hydroxylation of L-Orn were determined spectrophotometrically and are given with the kinetic parameters of several other FAD-dependent monooxygenases for comparison in Table 6.3.

**Table 6.3:** The kinetic parameters for EtcB-mediated hydroxylation of L-Orn in comparison with kinetic parameters of FAD-dependent monooxygenases involved in siderophore biosynthesis.

| enzyme | substrate | $K_M$ (mM) | $k_{cat}$ ($min^{-1}$) | $k_{cat}/K_m$ ($min^{-1} \cdot mM^{-1}$) |
|---|---|---|---|---|
| EtcB | L-Orn | 0.286 ± 0.035 | 19.6 ± 0.03 | 68.5 |
| CchB | L-Orn | 3.6 ± 0.58 | 17.4 ± 0.87 | 4.83 |
| VbsO | L-Orn | 0.305 ± 0.024 | 108 ± 2 | 354.1 |
| PvdA | L-Orn | 0.593 ± 0.012 | 26.4 ± 0.4 | 44.52 |
| IucD | L-Lys | 0.11 | n.d. | n.d. |

The value for $K_M$ with 0.286 ± 0.035 mM is in the range of the values determined for VbsO, PvdA and IucB but is lower than the $K_M$ of CchB, being indicative of a higher affinity of EtcB towards the cognate

substrate. Substrate turnover is also in the same range as CchB, PvdA and IucB but significantly lower than the value determined for VbsO. The catalytic efficiency is also lower than the efficiency observed for VbsO, but similar to the value determined for PvdA and higher than the value for CchB. In summary, EtcB represents a typical member of FAD-dependent monooxygenases involved in the δ-*N*-hydroxylation of L-Orn or L-Lys. Based on the results a mechanism of catalysis can be postulated for EtcB and Sace_1309 as members of the external FAD-dependent monooxygenases (Figure 6.13). In the first step of catalysis, the enzyme associated oxidized FAD cofactor is reduced under the consumption of NADPH, leading to the generation of $FADH_2$.[169] The reduced cofactor is subsequently converted into a highly reactive hydroperoxy-flavin species by molecular oxygen, which reacts with the substrate L-Orn to afford L-hOrn and hydroxy-flavin. Elimination of water from hydroxy-flavin the and release of L-hOrn results in the oxidized cofactor FAD and enables another catalytic cycle.[94]

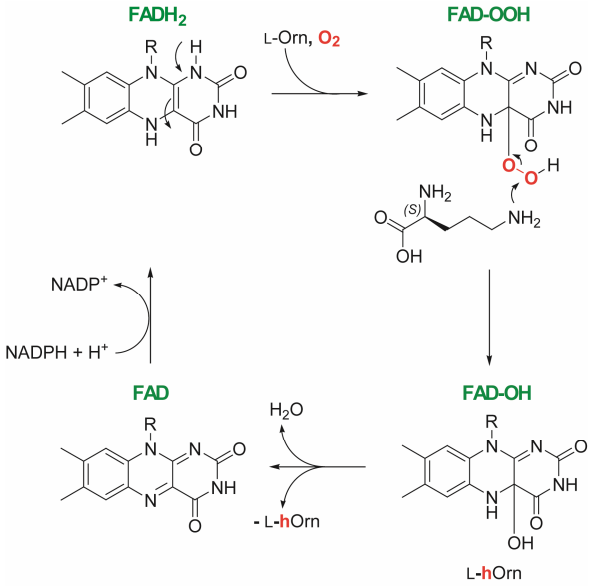



**Figure 6.13:** The proposed mechanism for δ-*N*-hydroxylation of L-Orn with detailed description of the resulting FAD species. In the first step the reduced FAD cofactor ($FADH_2$) is converted to a reactive hydroperoxy-flavin (FAD-OOH) by molecular oxygen. This reactive species is nucleophilically attacked by L-Orn to afford L-hOrn and hydroxy-flavin (FAD-OH). The elimination of $H_2O$ results in FAD, which is subsequently reduced to FADH under consumption of NADPH.

In summary, the biochemical characterization of EtcB and Sace_1309 established a biosynthetic route for the generation of δ-*N*-hydroxy-L-ornithine during erythrochelin biosynthesis *in vitro*. This building block represents a branching point in erythrochelin biosynthesis as it can be directly incorporated into the tetrapeptide siderophore at position 3 or be subjected to an additional tailoring step catalyzed by the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd (Chapter 6.3.2).

### 6.3.2 Characterization of Hydroxyornithine δ-*N*-Acetylation

#### 6.3.2.1 Bioinformatic Analysis of the Bifunctional Enzyme Mcd

Genes required for the biosynthesis of secondary metabolites are usually clustered within one region of the prokaryotic chromosome. As an example, all genes necessary for the assembly of the bacterial siderophore aerobactin, *iucA-iucD,* are located in a small region allowing coordinated regulation and expression.[99] In the aerobactin locus, several operator sequences for binding of the ferric uptake regulator (Fur) could be identified.[263] Analysis of the *etc* gene cluster revealed the lack of a gene encoding an *N*-acetyltransferase, responsible for the generation of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine, the residue incorporated into erythrochelin at positions 1 and 4. In order to identify a candidate gene, a genome-genome BLAST analysis was carried out with ACT and the genomes of *S. erythraea* and the aerobactin producing strain *Shigella flexneri* 5 str. 8401.[237] Analysis of the genome-genome comparison revealed *mcd* to display 33% sequential identity and 49% sequential similarity to the *N*-acetyltransferase IucB, catalyzing the acetylation of δ-*N*-hydroxy-L-lysine during the biosynthesis of the mixed citrate/hydroxamate-type siderophore aerobactin.[99] As seen in multiple sequence alignments, Mcd also shows homology to other putative *N*-acetyltransferases encoded in the biosynthetic operons for the production of mycobactin (MbtK, *Mycobacterium tuberculosis*) or vicibactin (VbsA, *Rhizobium leguminosarum*) (Figure 6.14).[162,260]

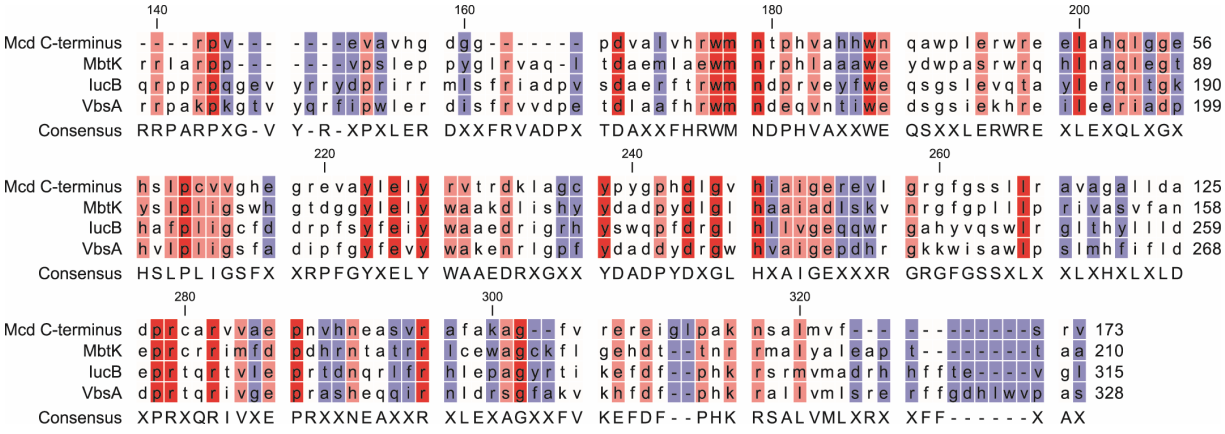
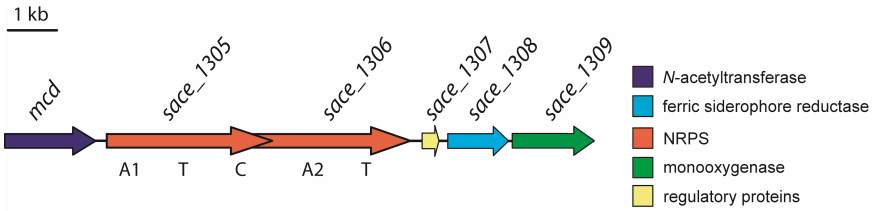



**Figure 6.14:** Multiple sequence alignments of the putative siderophore *N*-acetyltransferases Mcd (residues 243-417), MbtK, IucB and VbsA. Alignments were performed under application of a ClustalW algorithm. The degree of conservation is indicated by color: Red indicates complete and blue no agreement.

All *N*-acetyltransferases encoded in siderophore biosynthetic operons show sequential similarity to GCN5-related *N*-acetyltransferases (GNATs), involved in cellular processes such as the modification of aminoglycoside antibiotics or histone proteins in eukaryotes.[176,264] A subgroup of this enzyme class is represented by histone *N*-acetyltransferases (HATs), which mediate acetyltransfer to lysine residues in histones.[265] GNATs contain separate binding sites for the acyl group donor and the acceptor substrate, catalyzing the acyltransfer without the covalent binding of intermediates.[264] It was noted that within the GNAT superfamily, no conserved residue patterns could be identified that constitute an active site or are involved in the coordination of the acceptor substrate.[264] GNATs are mechanistically and structurally distinguished from PKS and FAS AT-domains, which catalyze the *S*-acylation of the ppan group of ACPs. In addition, PKS and FAS AT-domains bind their intermediate covalently and belong to the α/β-hydrolase superfamily, whereas GNATs display a specific fold.[264,266-267]

Mcd is encoded in the *nrps1* gene cluster which is located 2 mb away from the etc gene cluster in the core region of the *S. erythraea* genome (Figure 6.15).[215,238] Six genes constitute the *nrps1* gene cluster including two NRPSs with Sace_1305 showing an evident frameshift mutation rendering it non-functional. The product of *nrps1* is not known, but the ferric siderophore reductase Sace_1308 and the lysine/ornithine monooxygenase Sace_1309, which has been biochemically characterized *in vitro*, suggest the product of *nrps1* to be involved in cellular iron uptake.




| | |
|---|---|
| *mcd* | putative acetyl-CoA/*N*-acetyltransferase |
| *sace_1305* | nonribosomal peptide synthetase |
| *sace_1306* | nonribosomal peptide synthetase |
| *sace_1307* | MbtH-like protein |
| *sace_1308* | ferric siderophore reductase |
| *sace_1309* | putative peptide monooxygenase |

**Figure 6.15:** Schematic overview of the *nrps1* gene cluster. Putative functions of the proteins encoded within the operon are based on BLAST-analysis and are given in the figure. Apart from two NRPSs, a gene encoding a transcriptional regulator for secondary metabolism (*sace_1307*) and a ferric siderophore reductase (*sace_1308*) are found. The boundaries are determined by *mcd* coding for a putative *N*-acetyltransferase and *sace_1309*, encoding an ornithine monooxygenase.

Mcd has been shown to catalyze the biotin-independent decarboxylation of malonyl-CoA and additionally named EryM.[235] Disruption of *mcd* (*eryM*) gave rise to erythromycin nonproducing mutants and subsequent complementation studies confirmed Mcd to provide propionyl-CoA for the synthesis of

the PKS products erythromycin and the red pigment flaviolin via decarboxylation of methylmalonyl-CoA.[235] The *N*-terminal region of Mcd (residues 1-242) does not display sequential similarity to any protein deposited in public databases. Intriguingly, the *C*-terminal sequence of Mcd (residues 243-417) resembles the GNAT-domain found in the initiation module of CurA on a structural level.[238] CurA is involved in the biosynthesis of curacin A, which is assembled by a type I polyketide synthase in *Lyngbya majuscula*.[234,255] CurA contains an atypical *N*-terminal loading module, which also occurs in the biosynthetic pathway for onnamide. The CurA loading module is constituted of an *N*-terminal region of unknown function (~180 aa), designated adapter (AR)-domain which is not present in the OnnB loading module, a variable-length linker region and the $GNAT_L$-domain (Figure 6.16). The predicted function of such initiation modules is the loading of an acetyl group derived from acetyl-CoA onto the assembly line. It was shown that the $GNAT_L$-domain of CurA is a bifunctional enzyme that mediates malonyl-CoA decarboxylation to afford acetyl-CoA and subsequent acyltransfer to the ppan group of the adjacent ACP.[234] The three-dimensional structure of the CurA $GNAT_L$-domain cocrystallized with acetyl-CoA revealed unique substrate coordination and channelling mechanisms.

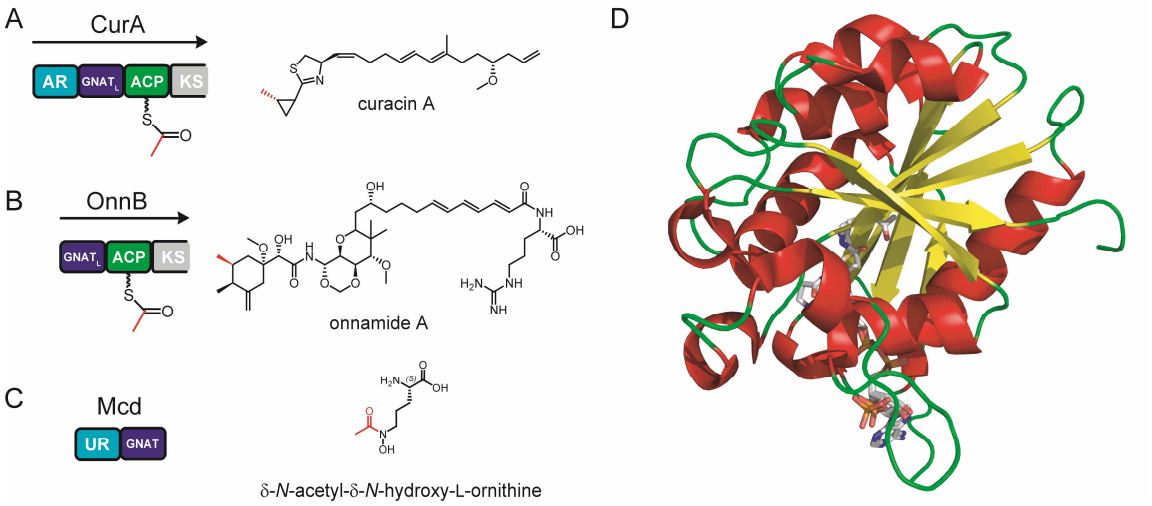



**Figure 6.16:** Initiation modules of PKSs containing $GNAT_L$-domains and Mcd together with the products assembled by the synthases and L-haOrn. A) Domain organization of CurA and chemical structure of curacin A. B) The onnamide loading module OnnB and onnamide A. C) The bifunctional malonyl-CoA decarboxylase/*N*-acetyltransferase Mcd and the final product L-haOrn. D) The three-dimensional structure of the CurA $GNAT_L$-domain (PDB code: 2REF) cocrystallized with acetyl-CoA. CurA exhibits the GNAT superfamily fold consisting of a central antiparallel β-sheet motif flanked by α-helices. Abbreviations: AR; adapter region-domain, UR; region of unknown function, KS; ketosynthase.

Multiple sequence alignments of the *C*-terminal residues of Mcd (243-417) and the *N*-terminal region of CurA (residues 218-440) could not identify a high degree of sequential similarity for the localization of the acetyl-CoA binding site or the residues involved in the decarboxylation reaction of malonyl-CoA. Within the crystal structure of the CurA $GNAT_L$-domain two residues (Thr355 and His389) could be identified at the junction of two tunnels separated by the aromatic side-chain of Trp249 (Figure 6.17).[234] Substitution of these residues with alanine resulted in decarboxylase activity deficient mutants, whereas acetyltransferase activity was still detectable. These residues could not be located within the sequences of Mcd or related GNAT-domain proteins associated with siderophore biosynthesis (Figure 6.14).

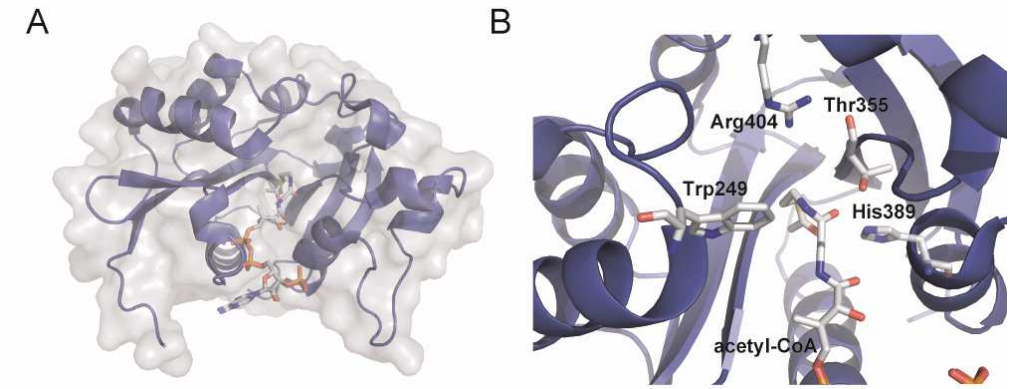



**Figure 6.17:** The three-dimensional structure of the CurA $GNAT_L$-domain (PDB code: 2REF). A) Surface representation of the GNAT-domain including the cosubstrate acetyl-CoA. B) Overview of the active site residues involved in the decarboxylation of malonyl-CoA. The substrate tunnels are separated by Trp249. Thr355 and His389 stabilize the enolate anion intermediate of the decarboxylation reaction. Arg404 is proposed to facilitate the binding of the deprotonated ACP ppan group.

#### 6.3.2.2 The Bifunctional Enzyme Mcd Mediates Decarboxylation and Acetylation

On the basis of the results obtained from bioinformatic analysis of *mcd* it was postulated that Mcd catalyzes the decarboxylation of malonyl-CoA to acetyl-CoA and subsequent acetyltransfer onto the δ-amino group of L-ornithine or δ-*N*-hydroxy-L-ornithine (Figure 6.18). In a study published recently, the REDIRECT method was employed for the generation of a *S. erythraea* Δ*mcd* mutant.[238]

Deletion of *mcd* resulted in an erythrochelin nonproducing strain and confirmed the essential role of Mcd in the biosynthesis of the siderophore. Complementation of the mutant strain with chemically synthesized L-haOrn restored erythrochelin production, proving Mcd to mediate δ-*N*-acetylation of free L-hOrn. On the basis of these results two biosynthetic routes for the generation of L-haOrn can be postulated. In the first model, hydroxylation of L-Orn, catalyzed by the FAD-dependent monooxygenases

EtcB or Sace_1309 precedes acetylation of L-hOrn by Mcd, whereas in the second model acetylation precedes hydroxylation (Figure 6.18).

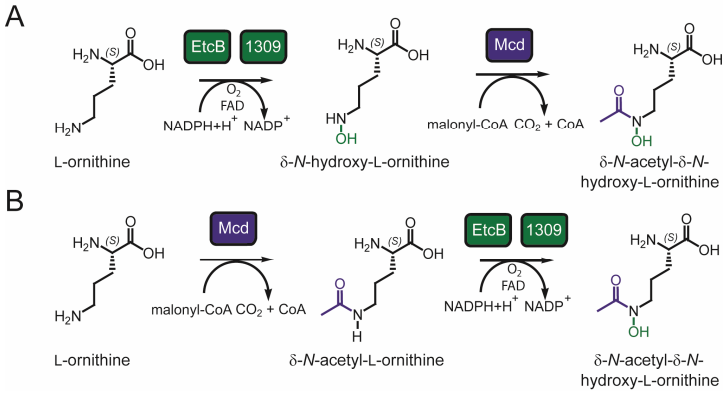



**Figure 6.18:** The mechanistic routes for the biosynthesis of L-haOrn. A) Hydroxylation of L-Orn by EtcB or Sace_1309 precedes acetylation. B) Acetylation of L-Orn by Mcd is followed by δ-*N*-hydroxylation.

To obtain further insights into the acetylation mechanism and timing, Mcd was produced recombinantly and characterized biochemically *in vitro*. It was shown that Mcd mediates the decarboxylation of malonyl-CoA and subsequent acetyltransfer to the δ-hydroxamino group of L-hOrn, demonstrating that hydroxylation of L-Orn precedes acetylation. Proving the resulting acetylated species to be δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine will require the chemical synthesis of an L-haOrn standard following established methods or the chemoenzymatic synthesis of preparative-scale amounts of the compound for NMR studies.[259]

Acetylation was exclusively limited to L-hOrn and confirmed Mcd to exhibit a high degree of substrate specificity. *In vitro* characterization of CurA also demonstrated a high degree of substrate specificity towards the acetyl group acceptor.[234] Substitution of the naturally occurring ACP ppan group by loading the CurA ACP with amino-CoA revealed only trace amounts of acetyl-NH-ACP. In contrast, Mcd is a GNAT-like protein catalyzing *N*-acetylation instead of *S*-acetylation. Time-course experiments showed that utilization of malonyl-CoA as the acetyl-donor resulted in only 35% substrate conversion after 240 min, whereas acetyl-CoA as the cognate acetyl group donor afforded complete substrate turnover after 90 min, proving the decarboxylation of malonyl-CoA and subsequent substrate channelling to be rate-limiting steps of this reaction. In contrast, the determination of kinetic parameters of *S*-acetylation by CurA using malonyl-CoA and acetyl-CoA revealed similar $k_{cat}$ values for both substrates. In addition, the $k_{cat}$ value for acetyl group transfer was 780-fold lower than the $k_{cat}$ for the decarboxylation of malonyl-CoA suggesting that decarboxylation and acetyltransfer are separated by conformational changes leading

to an effective coordination of the ACP ppan arm in the acceptor site.[234] The determination of kinetic parameters for Mcd-mediated acetylation of L-hOrn will grant further insights into this unique class of enzymes. In addition, a consecutive enzymatic synthesis of δ-*N*-acetyl-δ-*N*-hydroxy-L-ornithine was achieved by tandem action of a FAD-dependent monooxygenase, hydroxylating the δ-amino group of L-Orn, and the bifunctional malonyl-CoA decarboxylase/acetyltransferase Mcd. This coupled assay gave rise to the L-haOrn building block, which is subsequently recognized and activated by the first and the third A-domain of the synthetase EtcD *in vivo.*

### 6.4 Biosynthesis of Erythrochelin Requires NRPS Crosstalk

Biosynthesis of erythrochelin requires the crosstalk between distant NRPS gene clusters due to Mcd being encoded in a cluster 2 mb away from the *etc* locus.[238] The phenomenon of crosstalk between metabolic pathways is common in prokaryotes and has been observed for the malonyl-CoA transferase (MAT) of the *S. coelicolor* FAS, which is also proposed to be involved in the biosynthesis of the PKS-derived isochromanone pigment actinorhodin.[268] An additional example of one enzyme being shared by different biosynthetic pathways are the PPtases, catalyzing the conversion of FAS, PKS and NRPS *apo*-CPs to *holo*-CPs via phosphopantetheinylation in parallel. PPtases usually display a high degree of substrate tolerance towards the recognized CPs as in the case of the PPtase Sfp (*Bacillus subtilis*), whereas the SePptII PPtase (*S. erythraea*) is specific for the erythromycin PKS.[269]

Similar enzymes carrying out the same catalytic function in different pathways can complement each other upon deletion of one gene. This crosstalk has been described by mutational analysis of MbtH-like protein regulation of the coelichelin and CDA biosynthetic pathways in *S. coelicolor* A2(3).[168] MbtH-like proteins are small, highly conserved proteins of unknown function, assumed to be involved in the regulation of secondary metabolite biosynthesis and commonly found in NRPS biosynthetic gene clusters. Deletion of either *cdaX* or *cchX*, both encoding MbtH-like proteins, gave rise to mutants producing both compounds at wild-type levels. In contrast, the Δ*cdaX*Δ*cchX* double mutant failed to produce either of the compounds, confirming the necessity of the deleted genes for production of the corresponding metabolite.[168] In *S. erythraea*, MbtH-like proteins are encoded in the *nrps3*, *etc* and *nrps7* gene clusters.[215] Genetic deletion of *mcd* proved the gene encoding the bifunctional *N*-acetyltransferase to be required for erythrochelin assembly. As *mcd* is not located in the *etc* operon an unprecedented crosstalk between remote NRPS gene clusters has been uncovered. Intriguingly, Mcd was also shown to be essential for the biosynthesis of the macrolide polyketide antibiotic erythromycin, thus being involved in both PKS and NRPS pathways.[235]

### 6.5 Biosynthetic Model for Erythrochelin Assembly

On the basis of the results obtained herein, a model for erythrochelin biosynthesis by the tetramodular NRPS EtcD in combination with the FAD-dependent monooxygenases EtcB and Sace_1309 and the bifunctional tailoring enzyme Mcd was established (Figure 6.19). The irrevocable evidence for EtcD-mediated erythrochelin assembly resulted from targeted gene deletion of *etcD* followed by LCMS-analysis of culture supernatants. It was shown that deletion of *etcD* utilizing the RedET-based REDIRECT technique generated erythrochelin nonproducing mutants.[238] Eythrochelin biosynthesis by EtcD follows a linear enzymatic logic, in which the number of A-domains located within the template directly correlates with the number of amino acids found in the product. Initiation of erythrochelin assembly requires δ-*N*-hydroxylation of L-Orn by the flavin-dependent monooxygenase EtcB or Sace_1309, analogous to the CchB-catalyzed oxygenation of L-Orn during coelichelin biosynthesis.[98] L-hOrn itself represents a branching point in erythrochelin synthesis. This building block is either directly recognized by $A_3$ or further modified by means of δ-*N*-acetylation catalyzed by Mcd to afford L-haOrn which is recognized by $A_1$ and $A_4$. These results are consistent with bioinformatic analysis of EtcD adenylation domain specificity, resulting in the less accurate prediction of L-Arg as the substrate for both $A_1$ and $A_4$. Differences in the specificity determining residue pattern are likely to be the result of minimal structural differences between L-Arg and L-haOrn. When comparing the active site residues of $A_1$ and $A_4$, a high degree of identity (90%) is found, indicating L-haOrn as the common substrate. This model would exclude online δ-*N*-hydroxylation and δ-*N*-acetylation of the NRPS-bound substrates as seen in the hydroxylation of PCP-bound Glu in kutzneride biosynthesis.[102] Recent genetic studies investigating the biosynthesis of the L-haOrn residue corroborated this model.[238] Complementation of the Δ*mcd* mutant with chemically synthesized L-haOrn restored erythrochelin production, confirming the FAD-dependent monooxygenases EtcB and Sace_1309 to act in combination with Mcd for the biosynthesis of L-haOrn. This chemical complementation furthermore demonstrated EtcD $A_1$ and $A_4$ to recognize and activate the tailored L-haOrn as the cognate substrate.

Prior to incorporation of L-haOrn$_1$ into the growing peptide chain, the α-*N*-acetylation is likely to be carried out by the $C_1$-domain located at the *N*-terminus of EtcD, recognizing acetyl-CoA as the cognate substrate. This *in cis* modification of a PCP-bound substrate was shown to be adopted in the initiation reaction during surfactin biosynthesis with β-hydroxymyristoyl-CoA being the substrate for NRPS-catalyzed acyl transfer.[270] An alternative mechanism would arise from a lone-standing acetylated ACP being the cognate substrate donor similar to the initiation mechanism mediated by DptE and DptF during daptomycin biosynthesis (Figure 1.8).[106] Future work will focus on the reconstitution of the initiation of

erythrochelin biosynthesis *in vitro*, employing the *N*-terminal module 1 (C-$A_1$-PCP) of EtcD and acetyl-CoA as the acetyl-group donor. Epimerisation of the α-stereocenters of L-ac-haOrn$_1$ and L-Ser is mediated by the E-domains located in module 1 and 2, being in full agreement with the experimental determination of overall stereochemistry.

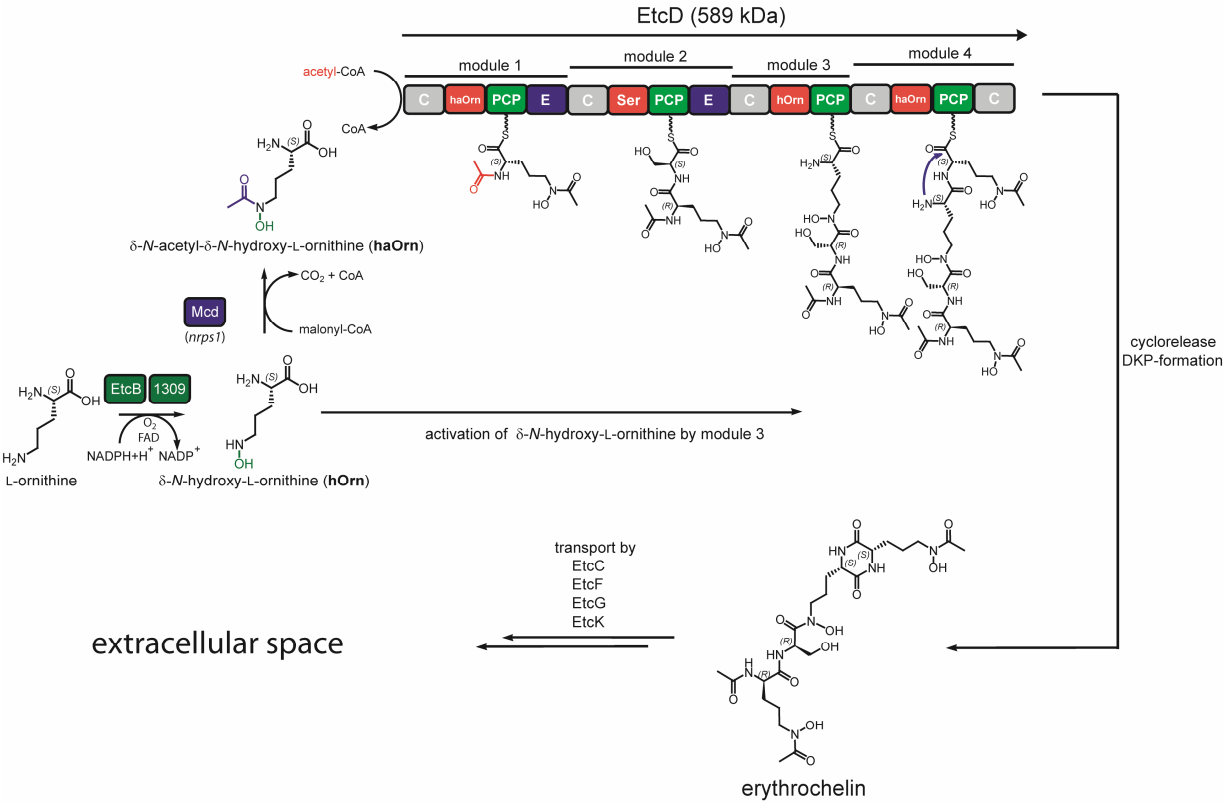



**Figure 6.19:** Biosynthesis of erythrochelin by the tetramodular nonribosomal peptide synthetase EtcD. δ-*N*-hydroxylation of L-ornithine is mediated by the peptide monooxygenases EtcB or Sace_1309. δ-*N*-acetylation of L-hOrn is carried out by the malonyl-CoA decarboxylase/*N*-acetyltransferase Mcd, encoded in the *nrps1* gene cluster. The *N*-terminal C-domain of EtcD catalyzes the α-*N*-acetylation of PCP-bound L-haOrn$_1$ *in cis*. Cyclorelease of the assembled tetrapeptide mediated by the *C*-terminal C-domain of EtcD results in the formation of a diketopiperazine (DKP) moiety. Erythrochelin is subsequently exported into the extracellular space to scavenge iron and is reimported into the cell by the uptake machinery consisting of EtcC/F/G.

The C-domain catalyzed condensation of the four unique building blocks follows a linear NRPS-assembly line logic. In the first step the $C_2$-domain catalyzes the nucleophilic attack of the Ser$_1$ α-amino group onto the $PCP_1$-bound ac-haOrn$_1$ resulting in a $PCP_2$-bound dipeptide. $C_3$-catalyzed isopeptide bond formation between the δ-amino group of L-hOrn$_3$ and the $PCP_2$-bound D-ac-haOrn$_1$-D-Ser$_2$ dipeptide results in the translocation of the tripeptide to $PCP_3$. A nucleophilic attack of the L-haOrn$_4$ α-amino group onto the

$PCP_3$-bound tripeptide thioester functionality gives the fully assembled tetrapeptide consisting of D-ac-haOrn$_1$-D-Ser$_2$-L-hOrn$_3$-L-haOrn$_4$. In contrast to erythrochelin assembly, the biosynthesis of the hydroxamate-type siderophores fuscachelin or coelichelin involves unorthodox use of the modular template including module-skipping mechanisms.[37-38]

The release of the assembled NRP is generally mediated by *C*-terminal TE-domains or reductase domains located in the termination module of the NRPS assembly line.[59,216] In contrast to this mechanism, the cyclorelease of erythrochelin through DKP-formation carried out by the *C*-terminal $C_5$-domain is proposed, catalyzing the intramolecular nucleophilic attack of the L-hOrn$_3$ α-amino group onto L-haOrn$_4$. Taking into account that the synthetases involved in the biosynthesis of the DKP-containing toxins thaxtomin (*Streptomyces acidiscabies*), gliotoxin (*Aspergillus fumigatus*) and fumitremorgin (*Aspergillus fumigatus*) also contain a *C*-terminal condensation domain, this C-domain catalyzed cyclorelease seems feasible (Figure 6.20 A).[61-62,271] Multiple sequence alignments of EtcD-$C_5$ with the *C*-terminal regions of TxtB, GliP and FtmA revealed EtcD-$C_5$ to contain a modified C3-core motif (Figure 6.20 B). The catalytically active second histidine residue of the C3-core motif (MHHxxxDG(WV)S) is missing with the core sequence being MHYLGDEWS. Future work will address the C-domain catalyzed cyclorelease by loading synthetic CoA-activated dipeptidyl-, tripeptidyl- and tetrapeptidyl substrates onto the recombinant EtcD PCP-$C_5$ didomain. In the course of this study the didomain could be produced recombinantly and the possibility to posttranslationally load synthetic CoA-substrates onto the PCP-domain was confirmed via fluoresceinyl-CoA phosphopantetheinylation assays (Figure S9).

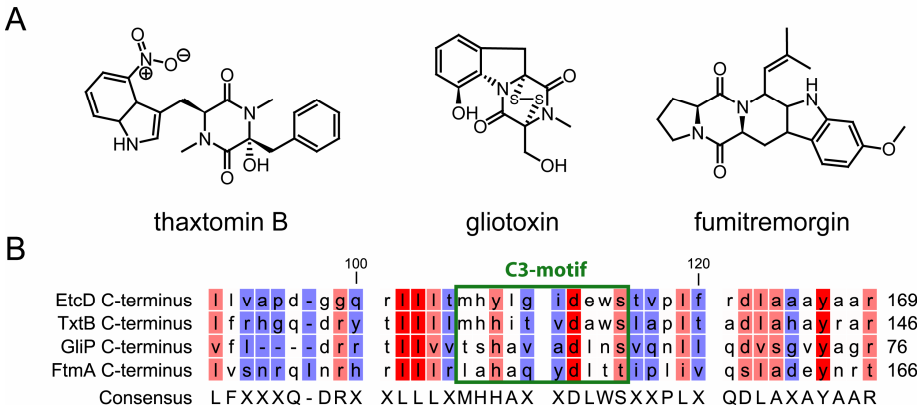



**Figure 6.20:** Multiple sequence alignments of *C*-terminal C-domains of NRPSs governing the biosynthesis of DKP-containing natural products and the corresponding chemical structures. A) The chemical structures of thaxtomin B, gliotoxin and fumitremorgin. B) Sequences aligned are EtcD, TxtB, GliP and FtmA, responsible for the assembly of erythrochelin, thaxtomin B, gliotoxin and fumitremorgin, respectively. The C3-core motif is accentuated in green. Alignments were performed under application of a ClustalW algorithm. The degree of conservation is indicated by color: Red indicates complete and blue no agreement.

After release from the synthetase, erythrochelin is exported into the extracellular space to scavenge iron. The reimport of *ferri*-erythrochelin is likely to be mediated by the FeuA homologue EtcC, responsible for periplasmic binding.[152] EtcC represents a membrane anchored periplasmic binding protein that in combination with EtcF, the ABC-transporter transmembrane component and EtcG, the corresponding ATP-binding component, actively reimports *ferri*-erythrochelin into the cell (Figure 6.21).[272] In order to investigate the recognition and binding of *ferri*-erythrochelin by EtcC, the protein was produced recombinantly omitting the *N*-terminal 21 residues, which are proposed to constitute the membrane anchor (Figure S10). Future work will aim to establish a structural and functional basis for the binding of *ferri*-erythrochelin by EtcC. In summary, the model for iron import by *S. erythraea* is as follows: I) Iron starvation leads to upregulation of the *etc* and *nrps1* gene clusters; II) Biosynthesis of the tailored amino acids L-hOrn and L-haOrn is carried out by EtcB, Sace_1309 and Mcd; III) The tetrapeptide siderophore is assembled by EtcD; IV) Export of the siderophore; V) Reimport of *ferri*-erythrochelin by EtcC/F/G.

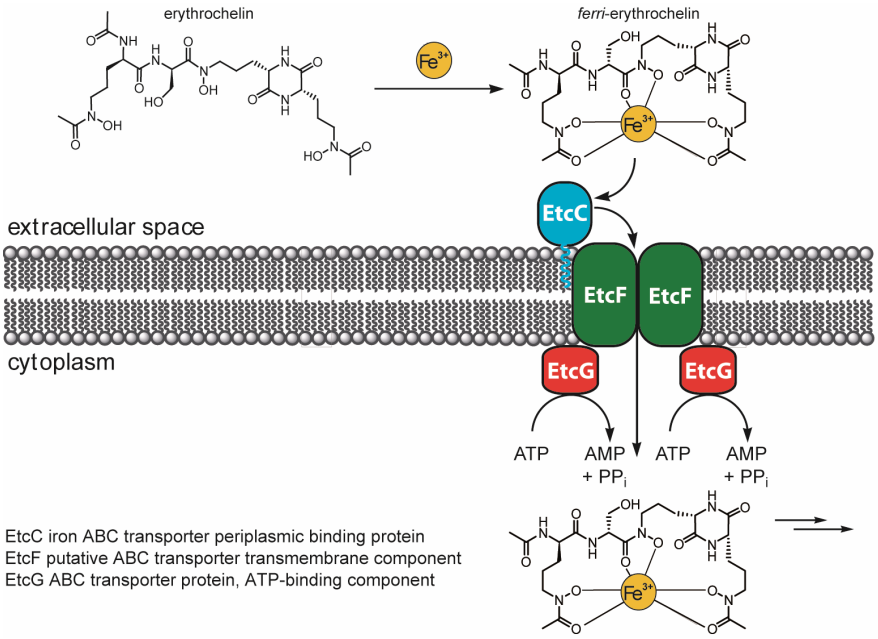



**Figure 6.21:** The proposed mechanism for iron uptake by *S. erythraea*. In response to iron starvation, erythrochelin is assembled and secreted into the extracellular space for the chelation of ferric iron. Loaded *ferri*-erythrochelin is recognized by the membrane-anchored periplasmic binding protein EtcC and channelled into the cytoplasm by the transport machinery consisting of EtcF and EtcG in an ATP-dependent manner.

# 7. References


1. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 70, 461-77 (2007).
2. Newman, D.J., Cragg, G.M. & Snader, K.M. The influence of natural products upon drug discovery. *Nat Prod Rep* 17, 215-34 (2000).
3. Faulkner, D.J. Marine pharmacology. *Antonie Van Leeuwenhoek* 77, 135-45 (2000).
4. Blunt, J.W., Copp, B.R., Munro, M.H., Northcote, P.T. & Prinsep, M.R. Marine natural products. *Nat Prod Rep* 27, 165-237 (2010).
5. Salomon, C.E., Magarvey, N.A. & Sherman, D.H. Merging the potential of microbial genetics with biological and chemical diversity: an even brighter future for marine natural product drug discovery. *Nat Prod Rep* 21, 105-21 (2004).
6. Knappe, T.A. et al. Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of Burkholderia thailandensis E264. *J Am Chem Soc* 130, 11446-54 (2008).
7. Schmidt, E.W. et al. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proc Natl Acad Sci U S A* 102, 7315-20 (2005).
8. Schwarzer, D., Finking, R. & Marahiel, M.A. Nonribosomal peptides: from genes to products. *Nat Prod Rep* 20, 275-87 (2003).
9. Du, L., Sanchez, C., Chen, M., Edwards, D.J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem Biol* 7, 623-42 (2000).
10. Kratzschmar, J., Krause, M. & Marahiel, M.A. Gramicidin S biosynthesis operon containing the structural genes grsA and grsB has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J Bacteriol* 171, 5422-9 (1989).
11. Zocher, R. et al. Biosynthesis of cyclosporin A: partial purification and properties of a multifunctional enzyme from Tolypocladium inflatum. *Biochemistry* 25, 550-3 (1986).
12. Enoch, D.A., Bygott, J.M., Daly, M.L. & Karas, J.A. Daptomycin. *J Infect* 55, 205-13 (2007).
13. Newman, D. & Cragg, G. Natural products in medicinal chemistry. *Bioorg Med Chem* 17, 2120 (2009).
14. Thomas, M.G., Chan, Y.A. & Ozanick, S.G. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob Agents Chemother* 47, 2823-30 (2003).
15. Nakano, M.M., Marahiel, M.A. & Zuber, P. Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in Bacillus subtilis. *J Bacteriol* 170, 5662-8 (1988).
16. Mittenhuber, G., Weckermann, R. & Marahiel, M.A. Gene cluster containing the genes for tyrocidine synthetases 1 and 2 from Bacillus brevis: evidence for an operon. *J Bacteriol* 171, 4881-7 (1989).
17. Guenzi, E., Galli, G., Grgurina, I., Gross, D.C. & Grandi, G. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J Biol Chem* 273, 32857-63 (1998).
18. Becker, J.E., Moore, R.E. & Moore, B.S. Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization. *Gene* 325, 35-42 (2004).
19. Tao, M. et al. The tallysomycin biosynthetic gene cluster from Streptoalloteichus hindustanus E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. *Mol Biosyst* 3, 60-74 (2007).
20. Robbel, L., Hoyer, K.M. & Marahiel, M.A. TioS T-TE--a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides. *Febs J* 276, 1641-53 (2009).
21. Robbel, L., Knappe, T.A., Linne, U., Xie, X. & Marahiel, M.A. Erythrochelin--a hydroxamate-type siderophore predicted from the genome of Saccharopolyspora erythraea. *Febs J* 277, 663-76 (2010).
22. Stein, T. et al. The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J Biol Chem* 271, 15428-35 (1996).
23. Konz, D. & Marahiel, M.A. How do peptide synthetases generate structural diversity? *Chem Biol* 6, R39-48 (1999).
24. Strieker, M., Tanovic, A. & Marahiel, M.A. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* 20, 234-40 (2010).
25. Walsh, C.T. & Fischbach, M.A. Natural products version 2.0: connecting genes to molecules. *J Am Chem Soc* 132, 2469-93 (2010).
26. Marahiel, M.A. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J Pept Sci* 15, 799-807 (2009).
27. Mootz, H.D., Schwarzer, D. & Marahiel, M.A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* 3, 490-504 (2002).
28. Tanovic, A., Samel, S.A., Essen, L.O. & Marahiel, M.A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* 321, 659-63 (2008).
29. Lambalot, R.H. et al. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* 3, 923-36 (1996).
30. Walsh, C.T. et al. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* 5, 525-34 (2001).
31. Samel, S.A., Marahiel, M.A. & Essen, L.O. How to tailor non-ribosomal peptide products--new clues about the structures and mechanisms of modifying enzymes. *Mol Biosyst* 4, 387-93 (2008).
32. von Dohren, H., Keller, U., Vater, J. & Zocher, R. Multifunctional Peptide Synthetases. *Chem Rev* 97, 2675-2706 (1997).
33. Schwarzer, D. & Marahiel, M.A. Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* 88, 93-101 (2001).
34. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347, 203-6 (1990).
35. Francklyn, C.S. DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression. *Biochemistry* 47, 11695-703 (2008).
36. Mootz, H.D. & Marahiel, M.A. The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* 179, 6843-50 (1997).
37. Dimise, E.J., Widboom, P.F. & Bruner, S.D. Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile Thermobifida fusca. *Proc Natl Acad Sci U S A* 105, 15311-6 (2008).

38. Lautru, S., Deeth, R.J., Bailey, L.M. & Challis, G.L. Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. *Nat Chem Biol* 1, 265-9 (2005).
39. Challis, G.L. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* 154, 1555-69 (2008).
40. Stachelhaus, T., Huser, A. & Marahiel, M.A. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem Biol* 3, 913-21 (1996).
41. Mootz, H.D., Finking, R. & Marahiel, M.A. 4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis. *J Biol Chem* 276, 37289-98 (2001).
42. Schmoock, G. et al. Functional cross-talk between fatty acid synthesis and nonribosomal peptide synthesis in quinoxaline antibiotic-producing streptomycetes. *J Biol Chem* 280, 4339-49 (2005).
43. Koglin, A. et al. Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* 312, 273-6 (2006).
44. Belshaw, P.J., Walsh, C.T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* 284, 486-9 (1999).
45. Linne, U. & Marahiel, M.A. Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* 39, 10439-47 (2000).
46. Luo, L. et al. Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* 41, 9184-96 (2002).
47. Ehmann, D.E., Trauger, J.W., Stachelhaus, T. & Walsh, C.T. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol* 7, 765-72 (2000).
48. Keating, T.A., Marshall, C.G., Walsh, C.T. & Keating, A.E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* 9, 522-6 (2002).
49. Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A. & Essen, L.O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 15, 781-92 (2007).
50. Balibar, C.J., Vaillancourt, F.H. & Walsh, C.T. Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol* 12, 1189-200 (2005).
51. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D.H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol* 7, 78 (2007).
52. Pazirandeh, M., Chirala, S.S., Huang, W.Y. & Wakil, S.J. Characterization of recombinant thioesterase and acyl carrier protein domains of chicken fatty acid synthase expressed in Escherichia coli. *J Biol Chem* 264, 18195-201 (1989).
53. Tseng, C.C. et al. Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* 41, 13350-9 (2002).
54. Hubbard, B.K. & Walsh, C.T. Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl* 42, 730-65 (2003).
55. Chiu, H.T. et al. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Natl Acad Sci U S A* 98, 8548-53 (2001).
56. Miller, D.A., Luo, L., Hillson, N., Keating, T.A. & Walsh, C.T. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem Biol* 9, 333-44 (2002).
57. Magarvey, N.A., Haltli, B., He, M., Greenstein, M. & Hucul, J.A. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrob Agents Chemother* 50, 2167-77 (2006).
58. Kessler, N., Schuhmann, H., Morneweg, S., Linne, U. & Marahiel, M.A. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J Biol Chem* 279, 7413-9 (2004).
59. Kopp, F., Mahlert, C., Grunewald, J. & Marahiel, M.A. Peptide macrocyclization: the reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine. *J Am Chem Soc* 128, 16478-9 (2006).
60. Weber, G. & Leitner, E. Disruption of the cyclosporin synthetase gene of Tolypocladium niveum. *Curr Genet* 26, 461-7 (1994).
61. Healy, F.G., Wach, M., Krasnoff, S.B., Gibson, D.M. & Loria, R. The txtAB genes of the plant pathogen Streptomyces acidiscabies encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol Microbiol* 38, 794-804 (2000).
62. Maiya, S., Grundmann, A., Li, S.M. & Turner, G. The fumitremorgin gene cluster of Aspergillus fumigatus: identification of a gene encoding brevianamide F synthetase. *Chembiochem* 7, 1062-9 (2006).
63. Debono, M. et al. Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J Antibiot (Tokyo)* 41, 1093-105 (1988).
64. Stein, D.B., Linne, U., Hahn, M. & Marahiel, M.A. Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis. *Chembiochem* 7, 1807-14 (2006).
65. Ansari, M.Z., Sharma, J., Gokhale, R.S. & Mohanty, D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinformatics* 9, 454 (2008).
66. de Crecy-Lagard, V. et al. Streptogramin B biosynthesis in Streptomyces pristinaespiralis and Streptomyces virginiae: molecular characterization of the last structural peptide synthetase gene. *Antimicrob Agents Chemother* 41, 1904-9 (1997).
67. Schauwecker, F., Pfennig, F., Grammel, N. & Keller, U. Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. *Chem Biol* 7, 287-97 (2000).
68. Perez Baz, J., Canedo, L.M., Fernandez Puentes, J.L. & Silva Elipe, M.V. Thiocoraline, a novel depsipeptide with antitumor activity produced by a marine Micromonospora. II. Physico-chemical properties and structure determination. *J Antibiot (Tokyo)* 50, 738-41 (1997).
69. Weber, G., Schorgendorfer, K., Schneider-Scherzer, E. & Leitner, E. The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* 26, 120-5 (1994).
70. Schoenafinger, G., Schracke, N., Linne, U. & Marahiel, M.A. Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J Am Chem Soc* 128, 7406-7 (2006).
71. Rouhiainen, L. et al. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90. *Mol Microbiol* 37, 156-67 (2000).

72. Schneider, T.L., Shen, B. & Walsh, C.T. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* 42, 9722-30 (2003).
73. Gehring, A.M., Mori, I., Perry, R.D. & Walsh, C.T. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis. *Biochemistry* 37, 11637-50 (1998).
74. Shen, B. et al. Cloning and characterization of the bleomycin biosynthetic gene cluster from Streptomyces verticillus ATCC15003. *J Nat Prod* 65, 422-31 (2002).
75. Konz, D., Klens, A., Schorgendorfer, K. & Marahiel, M.A. The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* 4, 927-37 (1997).
76. Reimmann, C., Serino, L., Beyeler, M. & Haas, D. Dihydroaeruginoic acid synthetase and pyochelin synthetase, products of the pchEF genes, are induced by extracellular pyochelin in Pseudomonas aeruginosa. *Microbiology* 144 ( Pt 11), 3135-48 (1998).
77. Reimmann, C. et al. Essential PchG-dependent reduction in pyochelin biosynthesis of Pseudomonas aeruginosa. *J Bacteriol* 183, 813-20 (2001).
78. Harle, J. & Bechthold, A. Chapter 12. The power of glycosyltransferases to generate bioactive natural compounds. *Methods Enzymol* 458, 309-33 (2009).
79. Chen, J. & Stubbe, J. Bleomycins: towards better therapeutics. *Nat Rev Cancer* 5, 102-12 (2005).
80. Wilkinson, B. & Micklefield, J. Chapter 14. Biosynthesis of nonribosomal peptide precursors. *Methods Enzymol* 458, 353-78 (2009).
81. van Wageningen, A.M. et al. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol* 5, 155-62 (1998).
82. He, H. et al. Mannopeptimycins, novel antibacterial glycopeptides from Streptomyces hygroscopicus, LL-AC98. *J Am Chem Soc* 124, 9729-36 (2002).
83. Cantoni, G.L. Biological methylation: selected aspects. *Annu Rev Biochem* 44, 435-51 (1975).
84. Schubert, H.L., Blumenthal, R.M. & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* 28, 329-35 (2003).
85. Hojati, Z. et al. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor. *Chem Biol* 9, 1175-87 (2002).
86. Mahlert, C., Kopp, F., Thirlway, J., Micklefield, J. & Marahiel, M.A. Stereospecific enzymatic transformation of alpha-ketoglutarate to (2S,3R)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J Am Chem Soc* 129, 12011-8 (2007).
87. Huang, Y.T. et al. In vitro characterization of enzymes involved in the synthesis of nonproteinogenic residue (2S,3S)-beta-methylphenylalanine in glycopeptide antibiotic mannopeptimycin. *Chembiochem* 10, 2480-7 (2009).
88. Vaillancourt, F.H., Yeh, E., Vosburg, D.A., Garneau-Tsodikova, S. & Walsh, C.T. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem Rev* 106, 3364-78 (2006).
89. O'Hagan, D., Schaffrath, C., Cobb, S.L., Hamilton, J.T. & Murphy, C.D. Biochemistry: biosynthesis of an organofluorine molecule. *Nature* 416, 279 (2002).
90. Rachid, S. et al. Molecular and biochemical studies of chondramide formation-highly cytotoxic natural products from Chondromyces crocatus Cm c5. *Chem Biol* 13, 667-81 (2006).
91. Vaillancourt, F.H., Yin, J. & Walsh, C.T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc Natl Acad Sci U S A* 102, 10111-6 (2005).
92. Broberg, A., Menkis, A. & Vasiliauskas, R. Kutznerides 1-4, depsipeptides from the actinomycete Kutzneria sp. 744 inhabiting mycorrhizal roots of Picea abies seedlings. *J Nat Prod* 69, 97-102 (2006).
93. Neumann, C.S. & Walsh, C.T. Biosynthesis of (-)-(1S,2R)-allocoronamic acyl thioester by an Fe(II)-dependent halogenase and a cyclopropane-forming flavoprotein. *J Am Chem Soc* 130, 14022-3 (2008).
94. van Berkel, W.J., Kamerbeek, N.M. & Fraaije, M.W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J Biotechnol* 124, 670-89 (2006).
95. Kopp, F., Linne, U., Oberthur, M. & Marahiel, M.A. Harnessing the chemical activation inherent to carrier protein-bound thioesters for the characterization of lipopeptide fatty acid tailoring enzymes. *J Am Chem Soc* 130, 2656-66 (2008).
96. Bruijnincx, P.C., van Koten, G. & Klein Gebbink, R.J. Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem Soc Rev* 37, 2716-44 (2008).
97. Costas, M., Mehn, M.P., Jensen, M.P. & Que, L., Jr. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem Rev* 104, 939-86 (2004).
98. Pohlmann, V. & Marahiel, M.A. Delta-amino group hydroxylation of L-ornithine during coelichelin biosynthesis. *Org Biomol Chem* 6, 1843-8 (2008).
99. de Lorenzo, V., Bindereif, A., Paw, B.H. & Neilands, J.B. Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in Escherichia coli K-12. *J Bacteriol* 165, 570-8 (1986).
100. Strieker, M., Kopp, F., Mahlert, C., Essen, L.O. & Marahiel, M.A. Mechanistic and structural basis of stereospecific Cbeta-hydroxylation in calcium-dependent antibiotic, a daptomycin-type lipopeptide. *ACS Chem Biol* 2, 187-96 (2007).
101. Helmetag, V., Samel, S.A., Thomas, M.G., Marahiel, M.A. & Essen, L.O. Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *Febs J* 276, 3669-82 (2009).
102. Strieker, M., Nolan, E.M., Walsh, C.T. & Marahiel, M.A. Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis. *J Am Chem Soc* 131, 13523-30 (2009).
103. Cosmina, P. et al. Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis. *Mol Microbiol* 8, 821-31 (1993).
104. Schultz, A.W. et al. Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J Am Chem Soc* 130, 4507-16 (2008).
105. Miao, V. et al. Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* 151, 1507-23 (2005).

106. Wittmann, M., Linne, U., Pohlmann, V. & Marahiel, M.A. Role of DptE and DptF in the lipidation reaction of daptomycin. *Febs J* 275, 5343-54 (2008).
107. Gehring, A.M., Mori, I. & Walsh, C.T. Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* 37, 2648-59 (1998).
108. Tomoda, H. et al. New cyclodepsipeptides, enniatins D, E and F produced by Fusarium sp. FO-1305. *J Antibiot (Tokyo)* 45, 1207-15 (1992).
109. Otsuka, H. & Shoji, J. The structure of Triostin C. *Tetrahedron* 21, 2931-8 (1965).
110. Juguet, M. et al. An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in Streptomyces ambofaciens. *Chem Biol* 16, 421-31 (2009).
111. Shaw-Reid, C.A. et al. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem Biol* 6, 385-400 (1999).
112. Hoyer, K.M., Mahlert, C. & Marahiel, M.A. The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization. *Chem Biol* 14, 13-22 (2007).
113. Nardini, M. & Dijkstra, B.W. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* 9, 732-7 (1999).
114. Bruner, S.D. et al. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* 10, 301-10 (2002).
115. Samel, S.A., Wagner, B., Marahiel, M.A. & Essen, L.O. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol* 359, 876-89 (2006).
116. Sieber, S.A. & Marahiel, M.A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* 105, 715-38 (2005).
117. Blow, D.M., Birktoft, J.J. & Hartley, B.S. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* 221, 337-40 (1969).
118. Kopp, F. & Marahiel, M.A. Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. *Curr Opin Biotechnol* 18, 513-20 (2007).
119. Kopp, F. & Marahiel, M.A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat Prod Rep* 24, 735-49 (2007).
120. Sieber, S.A., Tao, J., Walsh, C.T. & Marahiel, M.A. Peptidyl thiophenols as substrates for nonribosomal peptide cyclases. *Angew Chem Int Ed Engl* 43, 493-8 (2004).
121. Bycroft, B.W. et al. The total structure of viomycin, a tuberculostatic peptide antibiotic. *Experientia* 27, 501-3 (1971).
122. Wagner, B., Schumann, D., Linne, U., Koert, U. & Marahiel, M.A. Rational design of bacitracin A derivatives by incorporating natural product derived heterocycles. *J Am Chem Soc* 128, 10513-20 (2006).
123. Okada, H. et al. A new topoisomerase II inhibitor, BE-22179, produced by a streptomycete. I. Producing strain, fermentation, isolation and biological activity. *J Antibiot (Tokyo)* 47, 129-35 (1994).
124. Dawson, S., Malkinson, J.P., Paumier, D. & Searcey, M. Bisintercalator natural products with potential therapeutic applications: isolation, structure determination, synthetic and biological studies. *Nat Prod Rep* 24, 109-26 (2007).
125. Schaible, U.E. & Kaufmann, S.H. Iron and microbial infection. *Nat Rev Microbiol* 2, 946-53 (2004).
126. Waring, M.J. & Wakelin, L.P. Echinomycin: a bifunctional intercalating antibiotic. *Nature* 252, 653-7 (1974).
127. Wakelin, S.P. & Waring, M.J. The binding of echinomycin to deoxyribonucleic acid. *Biochem J* 157, 721-40 (1976).
128. Ughetto, G. et al. A comparison of the structure of echinomycin and triostin A complexed to a DNA fragment. *Nucleic Acids Res* 13, 2305-23 (1985).
129. Dell, A. et al. Structure revision of the antibiotic echinomycin. *J Am Chem Soc* 97, 2497-502 (1975).
130. Lee, J.S. & Waring, M.J. Bifunctional intercalation and sequence specificity in the binding of quinomycin and triostin antibiotics to deoxyribonucleic acid. *Biochem J* 173, 115-28 (1978).
131. Matson, J.A., Colson, K.L., Belofsky, G.N. & Bleiberg, B.B. Sandramycin, a novel antitumor antibiotic produced by a Nocardioides sp. II. Structure determination. *J Antibiot (Tokyo)* 46, 162-6 (1993).
132. Ohkuma, H. et al. BBM-928, a new antitumor antibiotic complex. I. Production, isolation, characterization and antitumor activity. *J Antibiot (Tokyo)* 33, 1087-97 (1980).
133. Nakaya, M. et al. Relative and absolute configuration of antitumor agent SW-163D. *Biosci Biotechnol Biochem* 71, 2969-76 (2007).
134. Negri, A. et al. Antitumor activity, X-ray crystal structure, and DNA binding properties of thiocoraline A, a natural bisintercalating thiodepsipeptide. *J Med Chem* 50, 3322-33 (2007).
135. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat Prod Rep* 27, 255-78 (2010).
136. Lombo, F. et al. Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two streptomyces species. *Chembiochem* 7, 366-76 (2006).
137. Kearns, K.D. & Hunter, M.D. Toxin-Producing Anabaena flos-aquae Induces Settling of Chlamydomonas reinhardtii, a Competing Motile Alga. *Microb Ecol* 42, 80-86 (2001).
138. Degrassi, G. et al. Plant growth-promoting Pseudomonas putida WCS358 produces and secretes four cyclic dipeptides: cross-talk with quorum sensing bacterial sensors. *Curr Microbiol* 45, 250-4 (2002).
139. Sheoran, A., King, A., Velasco, A., Pero, J.M. & Garneau-Tsodikova, S. Characterization of TioF, a tryptophan 2,3-dioxygenase involved in 3-hydroxyquinaldic acid formation during thiocoraline biosynthesis. *Mol Biosyst* 4, 622-8 (2008).
140. Biswas, T. et al. A new scaffold of an old protein fold ensures binding to the bisintercalator thiocoraline. *J Mol Biol* 397, 495-507 (2010).
141. Erba, E. et al. Mode of action of thiocoraline, a natural marine compound with anti-tumour activity. *Br J Cancer* 80, 971-80 (1999).
142. Romero, F. et al. Thiocoraline, a new depsipeptide with antitumor activity produced by a marine Micromonospora. I. Taxonomy, fermentation, isolation, and biological activities. *J Antibiot (Tokyo)* 50, 734-7 (1997).
143. Singh, R., Sharma, M., Joshi, P. & Rawat, D.S. Clinical status of anti-cancer agents derived from marine sources. *Anticancer Agents Med Chem* 8, 603-17 (2008).

144. Brandon, E.F. et al. In vitro characterization of the biotransformation of thiocoraline, a novel marine anti-cancer drug. *Invest New Drugs* 22, 241-51 (2004).
145. Bayo-Puxan, N. et al. Total solid-phase synthesis of the azathiocoraline class of symmetric bicyclic peptides. *Chemistry* 12, 9001-9 (2006).
146. Miethke, M. & Marahiel, M.A. Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev* 71, 413-51 (2007).
147. Winkelmann, G. & Drechsel, H. *Microbial Siderophores*, 200-246 (Wiley-VCH, Weinheim, 1997).
148. Carrano, C.J. & Raymond, K.N. Coordination chemistry of microbial iron transport compounds: rhodotorulic acid and iron uptake in Rhodotorula pilimanae. *J Bacteriol* 136, 69-74 (1978).
149. Tseng, C.F. et al. Bacterial siderophores: the solution stoichiometry and coordination of the Fe(III) complexes of pyochelin and related compounds. *J Biol Inorg Chem* 11, 419-32 (2006).
150. Winterberg, B. et al. Elucidation of the complete ferrichrome A biosynthetic pathway in Ustilago maydis. *Mol Microbiol* (2010).
151. Miethke, M. et al. Ferri-bacillibactin uptake and hydrolysis in Bacillus subtilis. *Mol Microbiol* 61, 1413-27 (2006).
152. Peuckert, F., Miethke, M., Albrecht, A.G., Essen, L.-O. & Marahiel, M.A. Structural Basis and Stereochemistry of Triscatecholate Siderophore Binding by FeuA. *Angew. Chem. Int. Ed.* 48(2009).
153. Andrews, S.C., Robinson, A.K. & Rodriguez-Quinones, F. Bacterial iron homeostasis. *FEMS Microbiol Rev* 27, 215-37 (2003).
154. Raymond, K.N., Dertz, E.A. & Kim, S.S. Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci U S A* 100, 3584-8 (2003).
155. May, J.J., Wendrich, T.M. & Marahiel, M.A. The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J Biol Chem* 276, 7209-17 (2001).
156. Bister, B. et al. The structure of salmochelins: C-glucosylated enterobactins of Salmonella enterica. *Biometals* 17, 471-81 (2004).
157. Dertz, E.A., Xu, J., Stintzi, A. & Raymond, K.N. Bacillibactin-mediated iron transport in Bacillus subtilis. *J Am Chem Soc* 128, 22-3 (2006).
158. Martinez, J.S. et al. Self-assembling amphiphilic siderophores from marine bacteria. *Science* 287, 1245-7 (2000).
159. Martinez, J.S. et al. Structure and membrane affinity of a suite of amphiphilic siderophores produced by a marine bacterium. *Proc Natl Acad Sci U S A* 100, 3754-9 (2003).
160. Meiwes, J. et al. Isolation and characterization of staphyloferrin A, a compound with siderophore activity from Staphylococcus hyicus DSM 20459. *FEMS Microbiol Lett* 55, 201-5 (1990).
161. Barbeau, K., Zhang, G., Live, D.H. & Butler, A. Petrobactin, a photoreactive siderophore produced by the oil-degrading marine bacterium Marinobacter hydrocarbonoclasticus. *J Am Chem Soc* 124, 378-9 (2002).
162. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H. & Walsh, C.T. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* 5, 631-45 (1998).
163. Haag, H. et al. Purification of yersiniabactin: a siderophore and possible virulence factor of Yersinia enterocolitica. *J Gen Microbiol* 139, 2159-65 (1993).
164. Bindereif, A. & Neilands, J.B. Cloning of the aerobactin-mediated iron assimilation system of plasmid ColV. *J Bacteriol* 153, 1111-3 (1983).
165. Kadi, N., Oves-Costales, D., Barona-Gomez, F. & Challis, G.L. A new family of ATP-dependent oligomerization-macrocyclization biocatalysts. *Nat Chem Biol* 3, 652-6 (2007).
166. Barona-Gomez, F., Wong, U., Giannakopulos, A.E., Derrick, P.J. & Challis, G.L. Identification of a cluster of genes that directs desferrioxamine biosynthesis in Streptomyces coelicolor M145. *J Am Chem Soc* 126, 16282-3 (2004).
167. Ledyard, K.M., Butler A. Structure of putrebactin, a new dihydroamate siderophore produced by Shewanella putrefaciens. *J Am Chem Soc* 2, 93-97 (1997).
168. Lautru, S., Oves-Costales, D., Pernodet, J.L. & Challis, G.L. MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in Streptomyces coelicolor M145. *Microbiology* 153, 1405-12 (2007).
169. Meneely, K.M. & Lamb, A.L. Biochemical characterization of a flavin adenine dinucleotide-dependent monooxygenase, ornithine hydroxylase from Pseudomonas aeruginosa, suggests a novel reaction mechanism. *Biochemistry* 46, 11930-7 (2007).
170. Gibson, F. & Magrath, D.I. The isolation and characterization of a hydroxamic acid (aerobactin) formed by Aerobacter aerogenes 62-I. *Biochim Biophys Acta* 192, 175-84 (1969).
171. Warner, P.J., Williams, P.H., Bindereif, A. & Neilands, J.B. ColV plasmid-specific aerobactin synthesis by invasive strains of Escherichia coli. *Infect Immun* 33, 540-5 (1981).
172. Carbonetti, N.H. & Williams, P.H. A cluster of five genes specifying the aerobactin iron uptake system of plasmid ColV-K30. *Infect Immun* 46, 7-12 (1984).
173. Thariath, A., Socha, D., Valvano, M.A. & Viswanatha, T. Construction and biochemical characterization of recombinant cytoplasmic forms of the IucD protein (lysine:N6-hydroxylase) encoded by the pColV-K30 aerobactin gene cluster. *J Bacteriol* 175, 589-96 (1993).
174. Coy, M., Paw, B.H., Bindereif, A. & Neilands, J.B. Isolation and properties of N epsilon-hydroxylysine:acetyl coenzyme A N epsilon-transacetylase from Escherichia coli pABN11. *Biochemistry* 25, 2485-9 (1986).
175. Kadi, N., Arbache, S., Song, L., Oves-Costales, D. & Challis, G.L. Identification of a gene cluster that directs putrebactin biosynthesis in Shewanella species: PubC catalyzes cyclodimerization of N-hydroxy-N-succinylputrescine. *J Am Chem Soc* 130, 10458-9 (2008).
176. Challis, G.L. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chembiochem* 6, 601-11 (2005).
177. Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A. & Walsh, C.T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* 407, 215-8 (2000).
178. Kohli, R.M., Walsh, C.T. & Burkart, M.D. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* 418, 658-61 (2002).
179. Sieber, S.A., Walsh, C.T. & Marahiel, M.A. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J Am Chem Soc* 125, 10862-6 (2003).

180. Kopp, F., Grunewald, J., Mahlert, C. & Marahiel, M.A. Chemoenzymatic design of acidic lipopeptide hybrids: new insights into the structure-activity relationship of daptomycin and A54145. *Biochemistry* 45, 10474-81 (2006).
181. Bode, H.B., Bethe, B., Hofs, R. & Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem* 3, 619-27 (2002).
182. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-80 (2005).
183. Donadio, S., Monciardini, P. & Sosio, M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* 24, 1073-109 (2007).
184. Ansari, M.Z., Yadav, G., Gokhale, R.S. & Mohanty, D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res* 32, W405-13 (2004).
185. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res* 33, 5799-808 (2005).
186. Hopwood, D.A. Forty years of genetics with Streptomyces: from in vivo through in vitro to in silico. *Microbiology* 145 ( Pt 9), 2183-202 (1999).
187. Bentley, S.D. et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417, 141-7 (2002).
188. Gust, B., Challis, G.L., Fowler, K., Kieser, T. & Chater, K.F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* 100, 1541-6 (2003).
189. Corre, C., Song, L., O'Rourke, S., Chater, K.F. & Challis, G.L. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. *Proc Natl Acad Sci U S A* 105, 17510-5 (2008).
190. Song, L. et al. Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by Streptomyces coelicolor genome mining. *J Am Chem Soc* 128, 14754-5 (2006).
191. Udwary, D.W. et al. Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proc Natl Acad Sci U S A* 104, 10376-81 (2007).
192. McClerren, A.L. et al. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc Natl Acad Sci U S A* 103, 17243-8 (2006).
193. Fortin, P.D., Walsh, C.T. & Magarvey, N.A. A transglutaminase homologue as a condensation catalyst in antibiotic assembly lines. *Nature* 448, 824-7 (2007).
194. Watanabe, K. et al. Escherichia coli allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product. *J Am Chem Soc* 131, 9347-53 (2009).
195. Watanabe, K., Oguri, H. & Oikawa, H. Diversification of echinomycin molecular structure by way of chemoenzymatic synthesis and heterologous expression of the engineered echinomycin biosynthetic pathway. *Curr Opin Chem Biol* 13, 189-96 (2009).
196. Gross, H. et al. The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem Biol* 14, 53-63 (2007).
197. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-5 (1970).
198. Sambrook, J. & Russel, D.W. *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2000).
199. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-54 (1976).
200. Berger, S. & Braun, S. *200 and More NMR Experiments. A Practical Course*, (Wiley-VCH, Weinheim, 2004).
201. Kessler, H., Schmieder, P., Köck, M. & Kurz, M. Improved Resolution in Proton-Detected Heteronuclear Long-Range Correlation. *J Magn Reson* 88, 615-618 (1990).
202. Cicero, D.O., Barbato, G. & Bazzo, R. Sensitivity enhancement of a two-dimensional experiment for the measurement of heteronuclear long-range coupling constants, by a new scheme of coherence selection by gradients. *J Magn Reson* 148, 209-13 (2001).
203. La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M. & Burkart, M.D. Manipulation of carrier proteins in antibiotic biosynthesis. *Chem Biol* 11, 195-201 (2004).
204. Patricelli, M.P., Giang, D.K., Stamp, L.M. & Burbaum, J.J. Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. *Proteomics* 1, 1067-71 (2001).
205. Schwyn, B. & Neilands, J.B. Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* 160, 47-56 (1987).
206. Bhushan, R. & Bruckner, H. Marfey's reagent for chiral amino acid analysis: a review. *Amino Acids* 27, 231-47 (2004).
207. Lin, Y. & Miller, M.J. Practical synthesis of hydroxamate-derived siderophore components by an indirect oxidation method and syntheses of a DIG-siderophore conjugate and a biotin-siderophore conjugate. *J Org Chem* 64, 7451-7458 (1999).
208. Cushman, D.W., Cheung, H.S., Sabo, E.F., Rubin, B. & Ondetti, M.A. Development of specific inhibitors of angiotensin I converting enzyme (kininase II). *Fed Proc* 38, 2778-82 (1979).
209. Rutherford, K. et al. Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944-5 (2000).
210. Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* 6, 493-505 (1999).
211. Carver, T. et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24, 2672-6 (2008).
212. Abbott, J.C., Aanensen, D.M. & Bentley, S.D. WebACT: an online genome comparison suite. *Methods Mol Biol* 395, 57-74 (2007).
213. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29, 2607-18 (2001).
214. Crotti, A.E.M. et al. The fragmentation mechanism of five-membered lactones by electrospray ionisation tandem mass spectrometry. *Int J Mass Spectrom* 232, 271-276 (2004).

215. Oliynyk, M. et al. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nat Biotechnol* 25, 447-53 (2007).
216. Kohli, R.M. & Walsh, C.T. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem Commun (Camb)*, 297-307 (2003).
217. Challis, G.L., Ravel, J. & Townsend, C.A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* 7, 211-24 (2000).
218. Tillett, D. et al. Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol* 7, 753-64 (2000).
219. Gehring, A.M., Bradley, K.A. & Walsh, C.T. Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* 36, 8495-503 (1997).
220. May, J.J., Kessler, N., Marahiel, M.A. & Stubbs, M.T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* 99, 12120-5 (2002).
221. McMorran, B.J., Kumara, H.M., Sullivan, K. & Lamont, I.L. Involvement of a transformylase enzyme in siderophore synthesis in Pseudomonas aeruginosa. *Microbiology* 147, 1517-24 (2001).
222. Takeuchi, N. et al. Mammalian mitochondrial methionyl-tRNA transformylase from bovine liver. Purification, characterization, and gene structure. *J Biol Chem* 273, 15085-90 (1998).
223. Pospiech, A., Bietenhader, J. & Schupp, T. Two multifunctional peptide synthetases and an O-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumour agent saframycin Mx1 from Myxococcus xanthus. *Microbiology* 142 ( Pt 4), 741-6 (1996).
224. Yu, S., Fiss, E. & Jacobs, W.R., Jr. Analysis of the exochelin locus in Mycobacterium smegmatis: biosynthesis genes have homology with genes of the peptide synthetase family. *J Bacteriol* 180, 4676-85 (1998).
225. Peano, C. et al. Complete gene expression profiling of Saccharopolyspora erythraea using GeneChip DNA microarrays. *Microb Cell Fact* 6, 37 (2007).
226. Gondry, M. et al. Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat Chem Biol* 5, 414-20 (2009).
227. Hu, J. & Miller, M.J. A new method for the synthesis of of Nε-acetyl-Nε-hydroxy-L-lysine, the iron-binding constituent of several Important siderophores. *J. Org. Chem* 59(1994).
228. Paul, M., Poyan Mehr, A. & Kreutz, R. Physiology of local renin-angiotensin systems. *Physiol Rev* 86, 747-803 (2006).
229. Umezawa, H. et al. Foroxymithine, a new inhibitor of angiotensin-converting enzyme, produced by actinomycetes. *J Antibiot (Tokyo)* 38, 1813-5 (1985).
230. Aoyagi, T. et al. Influence of angiotensin-converting enzyme inhibitor, foroxymithine, on dynamic equilibrium around the renin-angiotensin system in vivo. *J Appl Biochem* 7, 388-95 (1985).
231. Imoto, M. et al. Antitumor activity of erbstatin, a tyrosine protein kinase inhibitor. *Jpn J Cancer Res* 78, 329-32 (1987).
232. Martinez, J.L., Herrero, M. & de Lorenzo, V. The organization of intercistronic regions of the aerobactin operon of pColV-K30 may account for the differential expression of the iucABCD iutA genes. *J Mol Biol* 238, 288-93 (1994).
233. de Lorenzo, V. & Neilands, J.B. Characterization of iucA and iucC genes of the aerobactin system of plasmid ColV-K30 in Escherichia coli. *J Bacteriol* 167, 350-5 (1986).
234. Gu, L. et al. GNAT-like strategy for polyketide chain initiation. *Science* 318, 970-4 (2007).
235. Hsieh, Y.J. & Kolattukudy, P.E. Inhibition of erythromycin synthesis by disruption of malonyl-coenzyme A decarboxylase gene eryM in Saccharopolyspora erythraea. *J Bacteriol* 176, 714-24 (1994).
236. Hunaiti, A.R. & Kolattukudy, P.E. Malonyl-CoA decarboxylase from Streptomyces erythreus: purification, properties, and possible role in the production of erythromycin. *Arch Biochem Biophys* 229, 426-39 (1984).
237. Oves-Costales, D., Kadi, N. & Challis, G.L. The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chem Commun (Camb)*, 6530-41 (2009).
238. Lazos, O. et al. Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. *Chem Biol* 17, 160-73 (2010).
239. Grunewald, J., Sieber, S.A., Mahlert, C., Linne, U. & Marahiel, M.A. Synthesis and derivatization of daptomycin: a chemoenzymatic route to acidic lipopeptide antibiotics. *J Am Chem Soc* 126, 17025-31 (2004).
240. Koketsu, K., Oguri, H., Watanabe, K. & Oikawa, H. Enzymatic macrolactonization in the presence of DNA leading to triostin A analogs. *Chem Biol* 15, 818-28 (2008).
241. Watanabe, K. et al. Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli. *Nat Chem Biol* 2, 423-8 (2006).
242. Wang, C., Wesener, S.R., Zhang, H. & Cheng, Y.Q. An FAD-dependent pyridine nucleotide-disulfide oxidoreductase is involved in disulfide bond formation in FK228 anticancer depsipeptide. *Chem Biol* 16, 585-93 (2009).
243. Bürgi, H., Dunitz, J. & Shefter, E. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. *J. Am. Chem. Soc.* 95, 5065-5067 (1973).
244. Chatterjee, J., Mierke, D. & Kessler, H. N-methylated cyclic pentaalanine peptides as template structures. *J Am Chem Soc* 128, 15164-72 (2006).
245. Chatterjee, J., Mierke, D.F. & Kessler, H. Conformational preference and potential templates of N-methylated cyclic pentaalanine peptides. *Chemistry* 14, 1508-17 (2008).
246. Mahlert, C., Sieber, S.A., Grunewald, J. & Marahiel, M.A. Chemoenzymatic approach to enantiopure streptogramin B variants: characterization of stereoselective pristinamycin I cyclase from Streptomyces pristinaespiralis. *J Am Chem Soc* 127, 9571-80 (2005).
247. Yeh, E., Lin, H., Clugston, S.L., Kohli, R.M. & Walsh, C.T. Enhanced macrocyclizing activity of the thioesterase from tyrocidine synthetase in presence of nonionic detergent. *Chem Biol* 11, 1573-82 (2004).
248. Wagner, B., Sieber, S.A., Baumann, M. & Marahiel, M.A. Solvent engineering substantially enhances the chemoenzymatic production of surfactin. *Chembiochem* 7, 595-7 (2006).

249. Trauger, J.W., Kohli, R.M. & Walsh, C.T. Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry* 40, 7092-8 (2001).
250. Bu, X. et al. Synthesis of gramicidin S and its analogues via an on-resin macrolactamization assisted by a predisposed conformation of the linear precursors. *J Org Chem* 69, 2681-5 (2004).
251. Wadhwani, P., Afonin, S., Ieronimo, M., Buerck, J. & Ulrich, A.S. Optimized protocol for synthesis of cyclic gramicidin S: starting amino acid is key to high yield. *J Org Chem* 71, 55-61 (2006).
252. Cuesta-Seijo, J.A., Weiss, M.S. & Sheldrick, G.M. Serendipitous SAD phasing of an echinomycin-(ACGTACGT)2 bisintercalation complex. *Acta Crystallogr D Biol Crystallogr* 62, 417-24 (2006).
253. Garcia-Martin, F., Cruz, L.J., Rodriguez-Mias, R.A., Giralt, E. & Albericio, F. Design and synthesis of FAJANU: a de novo C(2) symmetric cyclopeptide family. *J Med Chem* 51, 3194-202 (2008).
254. Zerikly, M. & Challis, G.L. Strategies for the discovery of new natural products by genome mining. *Chembiochem* 10, 625-33 (2009).
255. Chang, Z. et al. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. *J Nat Prod* 67, 1356-67 (2004).
256. Diekmann, H. [Metabolic products of microorganisms. 81. Occurrence and structures of coprogen B and dimerum acid]. *Arch Mikrobiol* 73, 65-76 (1970).
257. Atkin, C.L. & Neilands, J.B. Rhodotorulic acid, a diketopiperazine dihydroxamic acid with growth-factor activity. I. Isolation and characterization. *Biochemistry* 7, 3734-9 (1968).
258. Dolence, E.K. & Miller, M.J. Synthesis of foroxymithine, a microbial fermentation product and angiotensin 1 converting enzyme inhibitor. *J. Org. Chem.* 56, 492-499 (1991).
259. Heemstra, J.R., Jr., Walsh, C.T. & Sattely, E.S. Enzymatic tailoring of ornithine in the biosynthesis of the Rhizobium cyclic trihydroxamate siderophore vicibactin. *J Am Chem Soc* 131, 15317-29 (2009).
260. Carter, R.A. et al. The vbs genes that direct synthesis of the siderophore vicibactin in Rhizobium leguminosarum: their expression in other genera requires ECF sigma factor RpoI. *Mol Microbiol* 44, 1153-66 (2002).
261. Lomri, N., Gu, Q. & Cashman, J.R. Molecular cloning of the flavin-containing monooxygenase (form II) cDNA from adult human liver. *Proc Natl Acad Sci U S A* 89, 1685-9 (1992).
262. Stehr, M. et al. A hydrophobic sequence motif common to N-hydroxylating enzymes. *Trends Biochem Sci* 23, 56-7 (1998).
263. de Lorenzo, V., Wee, S., Herrero, M. & Neilands, J.B. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor. *J Bacteriol* 169, 2624-30 (1987).
264. Dyda, F., Klein, D.C. & Hickman, A.B. GCN5-related N-acetyltransferases: a structural overview. *Annu Rev Biophys Biomol Struct* 29, 81-103 (2000).
265. Kouzarides, T. Histone acetylases and deacetylases in cell proliferation. *Curr Opin Genet Dev* 9, 40-8 (1999).
266. Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N.A. & Cane, D.E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu Rev Biochem* 76, 195-221 (2007).
267. Leibundgut, M., Maier, T., Jenni, S. & Ban, N. The multienzyme architecture of eukaryotic fatty acid synthases. *Curr Opin Struct Biol* 18, 714-25 (2008).
268. Revill, W.P., Bibb, M.J. & Hopwood, D.A. Purification of a malonyltransferase from Streptomyces coelicolor A3(2) and analysis of its genetic determinant. *J Bacteriol* 177, 3946-52 (1995).
269. Weissman, K.J., Hong, H., Oliynyk, M., Siskos, A.P. & Leadlay, P.F. Identification of a phosphopantetheinyl transferase for erythromycin biosynthesis in Saccharopolyspora erythraea. *Chembiochem* 5, 116-25 (2004).
270. Steller, S. et al. Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* 43, 11331-43 (2004).
271. Balibar, C.J. & Walsh, C.T. GliP, a multimodular nonribosomal peptide synthetase in Aspergillus fumigatus, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry* 45, 15029-38 (2006).
272. Hollenstein, K., Dawson, R.J. & Locher, K.P. Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol* 17, 412-8 (2007).

# 8. Supplementary Section

## 8.1 Supplementary Figures

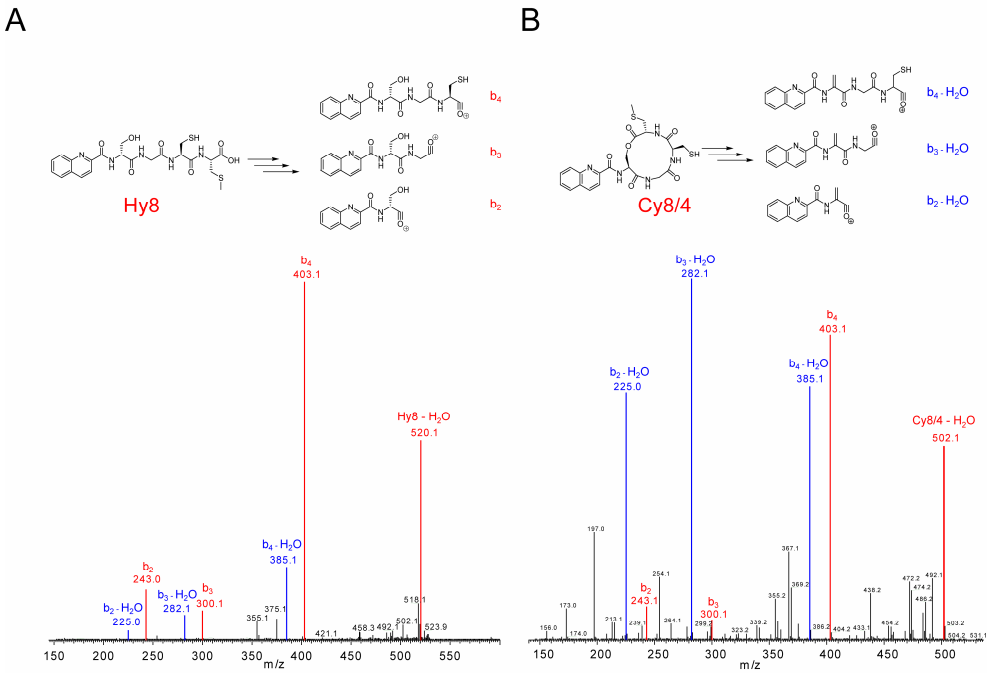



**Figure S1:** $MS^2$-spectra of the linear hydrolysis product Hy8 and macrolactone Cy8/4. A) $MS^2$-spectrum of the linear hydrolysis product Hy8 and the chemical structures of the resulting fragment ions. B) The fragmentation pattern of macrolactone Cy8/4 and the chemical structures of the resulting fragment ions. The fragmentation pattern differs from the spectrum of Hy8 due to the dominant formation of dehydro-alanine species ($b_{4/3/2}$-$H_2O$). The formation of dehydro-alanine species resulting from a loss of one molecule of water is characteristic for the fragmentation of lactones.[214]

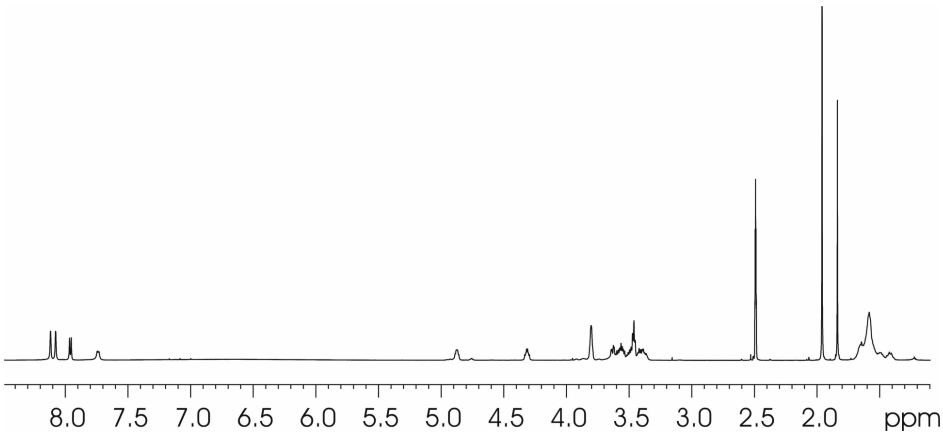



**Figure S2:** 1D $^1$H-NMR spectrum of erythrochelin in DMSO at 300K.

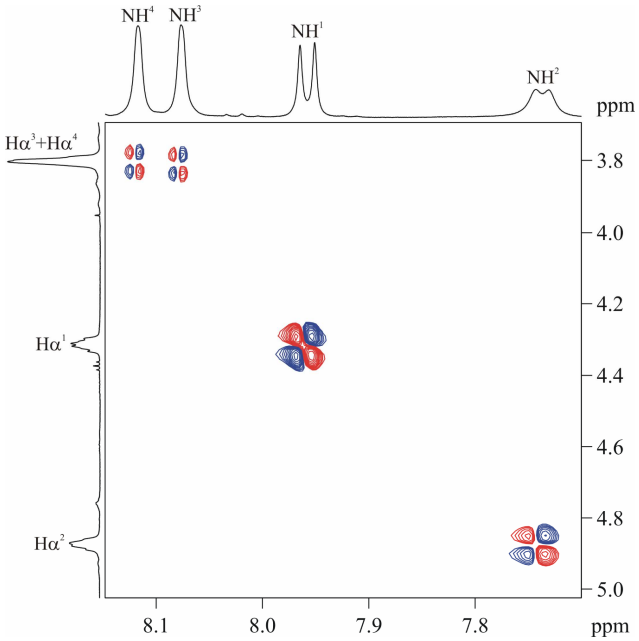



**Figure S3:** Section of the DQF-COSY spectrum of erythrochelin in DMSO at 300K.

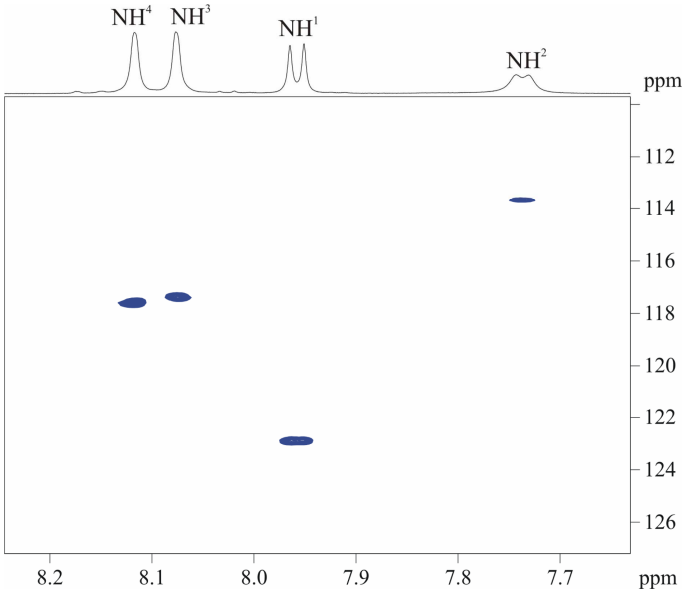



**Figure S4:** $^{1}H$-$^{15}N$ HSQC spectrum of erythrochelin in DMSO at 300K.

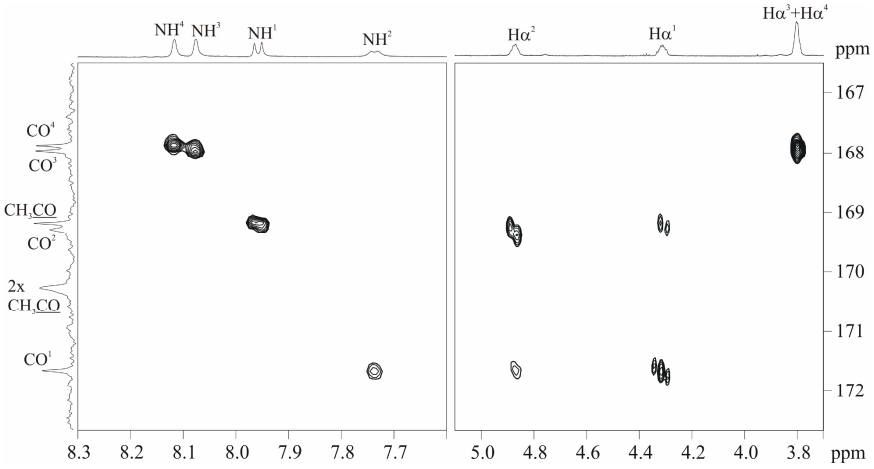



**Figure S5:** A section of the HMBC spectrum of erythrochelin in DMSO at 300K in the region of correlations between carbonyl atoms, amide protons and protons at the α-position.

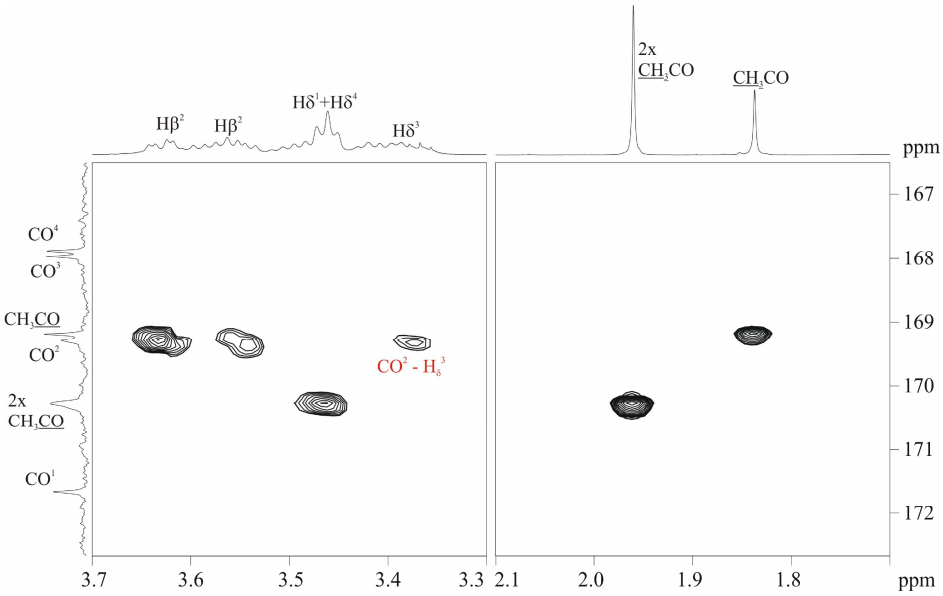



**Figure S6:** A section of the HMBC spectrum of erythrochelin in DMSO at 300K in the region of correlations between carbonyl groups and protons of amino acid sidechains.

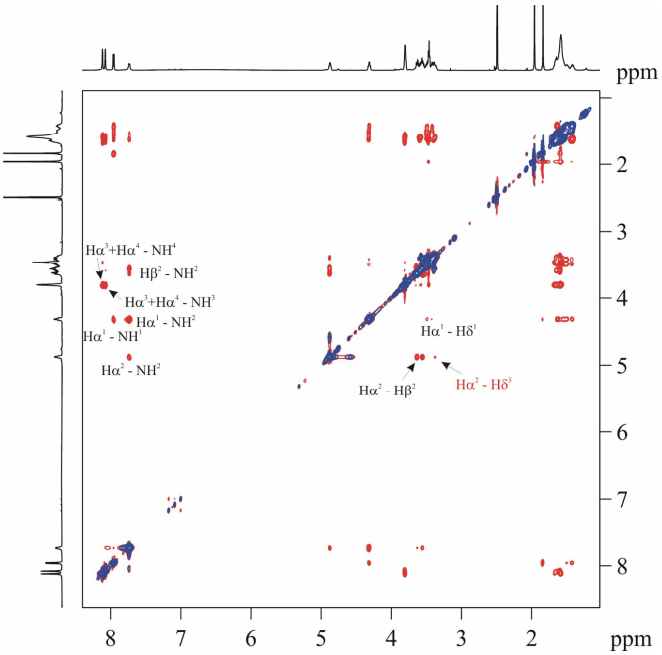



**Figure S7:** ROESY spectrum of erythrochelin in DMSO at 300K. The mixing time was 300 ms.

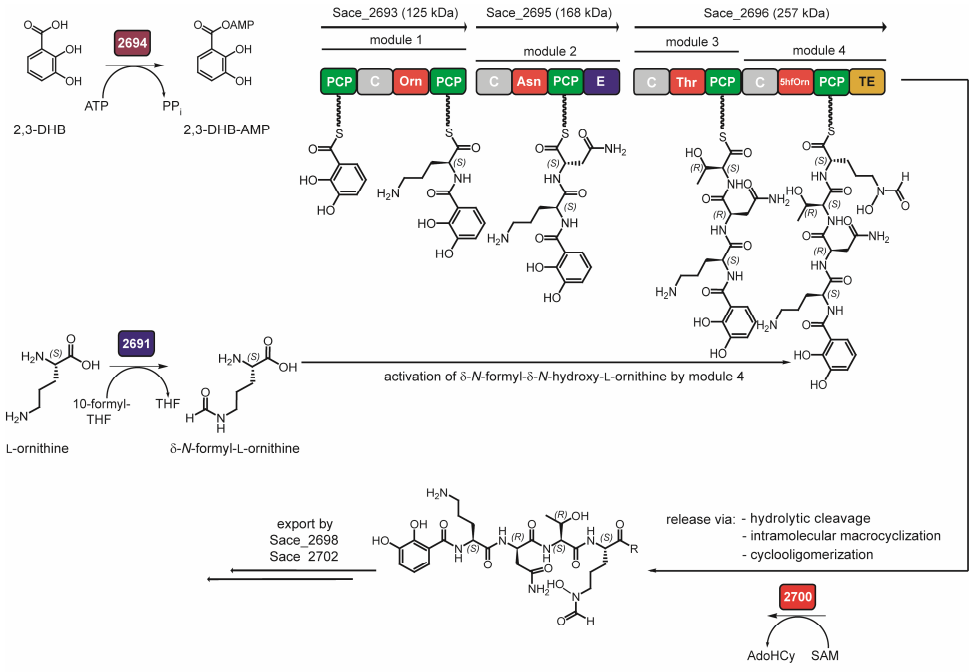



**Figure S8:** The proposed biosynthesis of the second siderophore by *nrps3* (Sace_ 2693, Sace_2695, Sace_2696). Initiation of siderophore biosynthesis is mediated by the lone-standing AMP-ligase Sace_2694, activating 2,3-DHB as aminoacyl-adenylate which is subsequently transferred onto the *N*-terminal PCP of Sace_2693. Additional tailoring steps involve the δ-*N*-formylation of L-Orn or L-hOrn to afford L-hfOrn. Release of the product is catalyzed by the *C*-terminal TE-domain of Sace_2696, giving rise to a mixed catecholate/hydroxamate-type siderophore.

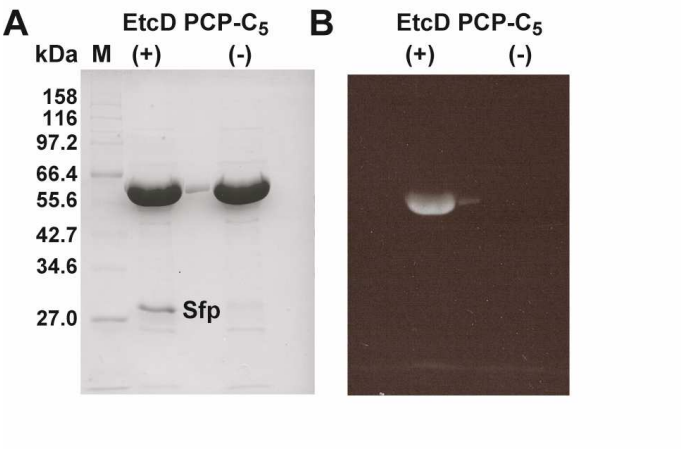



**Figure S9:** A) SDS-PAGE analysis of recombinant EtcD PCP-$C_5$ (58.3 kDa) incubated with fluoresceinyl-CoA and Sfp (+). (-) indicated the sample lacking Sfp. B) SDS-PAGE analysis of fluoresceinyl-labeled protein prior to Coomassie staining on an UV-screen ($\lambda$ = 312 nm). Labeling of the protein (+) indicates the possibility to posttranslationally load the didomain with CoA-activated peptidyl substrates. The control (-) lacked the PPtase Sfp. Protein Marker (M) was Broad Range Protein Marker P7702 (NEB).

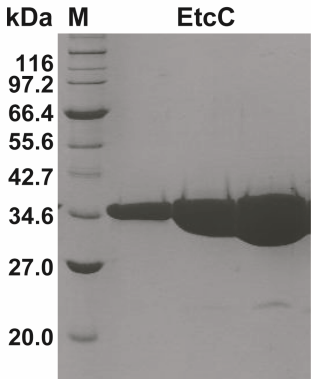



**Figure S10:** SDS-PAGE analysis of the recombinant periplasmic binding protein EtcC (36.1 kDa). Protein Marker (M) was Broad Range Protein Marker P7702 (NEB).

## 8.2 Supplementary Tables

**Table S1:** US FDA-approved NRPS-derived drugs. The brackets indicate the postulated system responsible for drug assembly.

| compound | system | class | indication | producing organism | semi-synthetic |
|---|---|---|---|---|---|
| **cyclic lipopeptides** | | | | | |
| daptomycin | NRPS | lipopeptide | antibacterial | *Streptomyces roseosporus* NRRL11379 | N |
| polymyxin B | (NRPS) | lipopeptide | antibacterial | *Bacillus polymyxa* | N |
| colistin | (NRPS) | lipopeptide | antibacterial | *Bacillus colistinus* | N |
| colistimethate | NRPS | lipopeptide | antibacterial | *Bacillus colistinus* | Y |
| caspofungin | (NRPS) | echinocandins | antifungal | *Aspergillus nidulans* | Y |
| micafungin | NRPS | echinocandins | antifungal | *Aspergillus nidulans* | Y |
| anidulafungin | NRPS | echinocandins | antifungal | *Aspergillus nidulans* | Y |
| **glycopeptides** | | | | | |
| vancomycin | NRPS | glycopeptide | antibacterial | *Amycolatopsis orientalis* | N |
| teichoplanin | PKS-NRPS | glycopeptide | antibacterial | *Actinoplanes teichomyceticus* | N |
| bleomycin | PKS-NRPS | glycopeptide | antineoplastic | *Streptomyces verticillus* ATCC 15003 | N |
| **cyclic peptides** | | | | | |
| bacitracin | NRPS | thiazolinepeptide | antibacterial | *Bacillus licheniformis* ATCC 10716 | N |
| quinopristin | NRPS | streptogramin | antibacterial | *Streptomyces pristinaespiralis* | Y |
| capreomycin | NRPS | aminoglycoside | antibacterial | *Streptomyces capreolus* ATCC 23892 | N |
| cyclosporine | NRPS | undecapeptide | immunosupressive | *Tolypocladium inflatum* | N |
| **linear peptides** | | | | | |
| gramicidin D | NRPS | oligopeptide | antibacterial | *Bacillus brevis* | N |
| **chromodepsipeptides** | | | | | |
| dactinomycin | NRPS | chromopeptide | antibacterial | *Streptomyces parvullus* | N |
| **tetrahydroisoquinolines** | | | | | |
| ecteinascidin ET-743 | NRPS | tetrahydroisoquinoline | antineoplastic | *Streptomyces lavendulae* | N |
| **β-lactam antibiotics** | | | | | |
| cephalosporin | NRPS | β-lactam | antibacterial | *Cephalosporium acremonium* | N |
| penicillin V/G | NRPS | β-lactam | antibacterial | *Penicillium sp.* | Y |
| ampicillin | NRPS | β-lactam | antibacterial | fungi | Y |
| carbenicillin | NRPS | β-lactam | antibacterial | fungi | Y |
| cefmetazole | NRPS | β-lactam | antibacterial | fungi | Y |
| cephamycin | NRPS | β-lactam | antibacterial | fungi | Y |
| loracarbef | NRPS | β-lactam | antibacterial | fungi | Y |
| cefotiam | NRPS | β-lactam | antibacterial | fungi | Y |
| flucloxacillin | NRPS | β-lactam | antibacterial | fungi | Y |

Source: U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Department of Health and Human Services, 25.03.2010

**Table S2:** Bioinformatic overview of the gene cluster responsible for erythrochelin biosynthesis and transport. The length of each gene is given in basepairs, whereas the length of the gene product is given in the number of amino acids. The proposed functions of the encoded proteins and the corresponding homology to related proteins are based on BLAST-analysis.

| gene | length (bp) | length (aa) | proposed function | sequence similarity, organism | identity / similarity (%) |
|---|---|---|---|---|---|
| *etcA* | 894 | 297 | LysR family transcriptional regulator | SALBJ_010100018803, *Streptomyces albus* J1074 | 44/56 |
| *etcB* | 1326 | 441 | putative peptide monooxygenase | SCO0498, *Streptomyces coelicolor* A3(2) | 69/82 |
| *etcC* | 987 | 328 | iron ABC transporter periplasmic-binding protein | AMIRDRAFT_47480, *Actinosynnema mirum* DSM 43827 | 51/73 |
| *etcD* | 16290 | 5429 | putative nonribosomal peptide synthetase | RHA1_RO04715, *Rhodococcus sp.* RHA1 | 50/63 |
| *etcE* | 204 | 67 | MbtH protein | SCO3218, *Streptomyces coelicolor* A3(2) | 72/83 |
| *etcF* | 1758 | 585 | putative ABC transporter transmembrane component | SROSN1_010100030367, *Streptomyces roseosporus* NRRL 15998 | 59/72 |
| *etcG* | 1638 | 545 | ABC transporter protein, ATP-binding component | SSDG_00035, *Streptomyces pristinaespiralis* ATCC 2586 | 60/72 |
| *etcH* | 666 | 221 | IclR-type transcriptional regulator | RHA1_RO05075, *Rhodococcus sp.* RHA1 | 52/66 |
| *etcI* | 1311 | 436 | CoA-transferase | ROP_51380, *Rhodococcus opacus* B4 | 73/82 |
| *etcJ* | 867 | 288 | hydroxymethylglutaryl-CoA lyase | RHA1_RO05078, *Rhodococcus jostii* RHA1 | 60/73 |
| *etcK* | 1293 | 430 | dicarboxylate carrier protein | RHA1_RO05079, *Rhodococcus jostii* RHA1 | 55/73 |

**Table S3:** $^{1}$H chemical shifts of erythrochelin in DMSO ($\delta_{1H}$ = 2.49 ppm) at 300 K.

| $^{1}$H (ppm) | NH | Hα | Hβ | Hγ | Hδ | Me |
|---|---|---|---|---|---|---|
| acetyl | | | | | | 1.838 |
| D-haOrn$_1$ | 7.958 (d: 8.3 Hz) | 4.318 | 1.422; 1.624 | 1.50; 1.54 | 3.416; 3.488 | 1.960 |
| D-Ser$_2$ | 7.737 (d: 7.4 Hz) | 4.875 | 3.552; 3.626 | / | / | / |
| L-hOrn$_3$ | 8.077 (s) | 3.811 | 1.594 | 1.59 | 3.38; 3.58 | |
| L-haOrn$_4$ | 8.117 (s) | 3.803 | 1.651 | 1.58 | 3.46 | 1.960 |

**Table S4:** $^{13}$C chemical shifts of erythrochelin in DMSO ($\delta_{13C}$ = 39.5 ppm) at 300 K.

| $^{13}$C (ppm) | Cα | Cβ | Cγ | Cδ | C=O | NC=O | Me |
|---|---|---|---|---|---|---|---|
| acetyl | | | | | | 169.20 | 22.51 |
| D-haOrn$_1$ | 52.13 | 29.45 | 23.09 | 46.67 | 171.67 | 170.28 | 20.31 |
| D-Ser$_2$ | 52.13 | 60.86 | / | / | 169.30 | / | / |
| L-hOrn$_3$ | 53.78 | 30.09 | 21.76 | 47.10 | 167.97 | / | / |
| L-haOrn$_4$ | 53.57 | 30.28 | 22.08 | 46.81 | 167.88 | 170.28 | 20.31 |

**Table S5:** $^{15}$N chemical shifts[a] of erythrochelin in DMSO at 300 K.

| $^{15}$N (ppm) | HN |
|---|---|
| D-haOrn$_1$ | 122.86 |
| D-Ser$_2$ | 113.60 |
| L-hOrn$_3$ | 117.35 |
| L-haOrn$_4$ | 117.60 |

[a]referenced to $^{15}$N of urea at 77.0 ppm.

**Table S6:** The observed NOE contacts of erythrochelin in DMSO at 300K.

| cross peaks | intensity |
|---|---|
| NH (D-haOrn$_1$) – Hα (D-haOrn$_1$) | strong (s) |
| NH (D-Ser$_2$) – Hα (D-haOrn$_1$) | s |
| NH (D-Ser$_2$) – Hα (D-Ser$_2$) | s |
| NH (D-Ser$_2$) – Hβ (D-Ser$_2$) | medium (m) |
| NH (L-hOrn$_3$) – Hα (L-hOrn$_3$/L-haOrn$_4$) | s |
| NH (L-haOrn$_4$) – Hα (L-hOrn$_3$/L-haOrn$_4$) | s |
| NH (D-haOrn$_1$) – CH$_3$ (Ac) | s |
| NH (D-haOrn$_1$) – Hβ (D-haOrn$_1$) | m |
| NH (D-Ser$_2$) – Hβ (D-haOrn$_1$) | weak (w) |
| NH (L-hOrn$_3$) – Hβ (L-hOrn$_3$) | m |
| NH (L-haOrn$_4$) – Hβ (L-haOrn$_4$) | m |
| Hα (D-haOrn$_1$) – Hβ (D-haOrn$_1$) | m |
| Hα (D-haOrn$_1$) – Hδ (D-haOrn$_1$) | w |
| Hα (D-Ser$_2$) – Hβ (D-Ser$_2$) | m |
| Hα (D-Ser$_2$) – Hδ (L-hOrn$_3$) | w |
| Hα (L-hOrn$_3$) – Hβ (L-hOrn$_3$) | s |
| Hα (L-haOrn$_4$) – Hβ (L-haOrn$_4$) | s |
| CH$_3$ (δ-*N*-ac-D-haOrn$_1$) – Hδ (D-haOrn$_1$) | w |
| CH$_3$ (δ-*N*-ac-D-haOrn$_4$) – Hδ (D-haOrn$_4$) | w |

**Table S7:** The observed long-range $^{1}$H-$^{13}$C correlations in erythrochelin in DMSO at 300K.

| cross peaks | coupling constants | intensity |
|---|---|---|
| CH$_3$(α-*N*-ac-L-haOrn$_1$) – C=O (α-*N*-ac) | $^{2}J_{CH}$ | s |
| C=O (α-*N*-ac) – NH (D-haOrn$_1$) | $^{2}J_{CH}$ | s |
| NH (D-haOrn$_1$) – Cα (D-haOrn$_1$) | $^{2}J_{CH}$ | s |
| NH (D-haOrn$_1$) – Cβ (D-haOrn$_1$) | $^{3}J_{CH}$ | s |
| NC=O (D-haOrn$_1$) – CH$_3$ (D-haOrn$_1$) | $^{2}J_{CH}$ | s |
| NC=O (D-haOrn$_1$) – Hδ (D-haOrn$_1$) | $^{3}J_{CH}$ | s |
| Hα (D-haOrn$_1$) – C=O (Ac) | $^{3}J_{CH}$ | w |
| Hα (D-haOrn$_1$) – C=O (D-haOrn$_1$) | $^{2}J_{CH}$ | s |
| C=O (D-haOrn$_1$) – NH (D-Ser$_2$) | $^{2}J_{CH}$ | w |
| Hα(D-Ser$_2$) – C=O (D-haOrn$_1$) | $^{3}J_{CH}$ | w |
| Hα (D-Ser$_2$) – C=O (D-Ser$_2$) | $^{2}J_{CH}$ | s |
| Hβ (D-Ser$_2$) – C=O (D-Ser$_2$) | $^{3}J_{CH}$ | s |
| Hα (L-hOrn$_3$) – C=O (L-hOrn$_3$) | $^{2}J_{CH}$ | s |
| C=O (D-Ser$_2$) – Hδ (L-hOrn$_3$) | $^{3}J_{CH}$ | w |
| Hα (L-haOrn$_4$) – C=O (L-haOrn$_4$) | $^{2}J_{CH}$ | s |
| NH (L-hOrn$_3$) – Cα (L-hOrn$_3$) | $^{2}J_{CH}$ | s |
| NH (haOrn$_4$) – Cα (haOrn$_4$) | $^{2}J_{CH}$ | s |
| C=O (δ-*N*-ac-L-haOrn$_4$) – CH$_3$ (L-haOrn$_4$) | $^{2}J_{CH}$ | s |
| C=O (δ-*N*-ac-L-haOrn$_4$) – Hδ (L-haOrn$_4$) | $^{3}J_{CH}$ | s |

## Acknowledgements

First of all, I would like to thank Prof. Dr. M. A. Marahiel for being an excellent supervisor and mentor over the last years. His open-mindness towards new ideas was a constant source of motivation and together with his continuous willingness for scientific discussions paved the way for the development of new concepts. Discussions with Prof. Dr. M. A. Marahiel going beyond science also contributed to the unique atmosphere within Marahiel group that I have experienced and enjoyed throughout the years. I also greatly acknowledge Prof. Dr. M. A. Marahiel for granting me the possibility to continue the projects developed from this thesis.
I would also like to thank him for giving me the opportunity to attend numerous conferences and congresses for the presentation of our data, such as the „PKS/NRPS-meeting“ (Hirschegg, Austria) and the “ChemBionics: Prospects of biohybrid molecules congress” (Rauischholzhausen, Germany). Especially, I would like to thank him for giving me the opportunity to attend the “Microbial genomics and secondary metabolites workshop” (Mediterranean Institute for Life Sciences, Split, Croatia) in 2007 under administration of Prof. Dr. D. Hopwood and Prof. Dr. J. Davies, the “PKS/NRPS-meeting” (Brown University, Providence, USA) and the “Biology of Streptomycetes congress” (Münster, Germany), which all have been outstanding experiences and expanded my scientific and personal horizon.

I gratefully thank Prof. Dr. L.-O. Essen for reviewing this thesis and being part of the thesis comitee.

Prof. Dr. A. Geyer and Prof. Dr. W. Buckel are gratefully appreciated for being part of the thesis comitee.

I would also like to thank Dr. Uwe Linne for his excellent support concerning bioanalytical questions. His profound expertise has clearly accelerated structure elucidation of erythrochelin and numerous MS-based analyses.

Dr. Xiulan Xie (NMR-department, Philipps-University Marburg) is gratefully acknowledged for conducting extensive and challenging NMR analyses. Her perseverance paved the way for structure elucidation of erythrochelin.

I acknowledge the whole Marahiel group for the nice working atmosphere, their support, a lot of fruitful discussions about and beyond science and the great time in the lab. Additionally, I would like to appreciate the hours outside the laboratory, which were always a constant source of relaxation and excitement at the same time. It was really a great time being part of the Marahiel group. In particular, I would like to thank current and former colleagues of Lab 4710 and the group: Dr. Matthias Strieker, Dr. Katharina Hoyer, Dr. Georg Schönafinger, Florian Peuckert, Alexander Albrecht and Mattia Bosello for a great time in the laboratory and for inspiring discussions.

Especially, I would like to thank the “Nightshift” consisting of Dr. Thomas Knappe and Alan Tanovic for the pleasant time in the laboratory even late in the evening.

I would like to thank the technician Antje Schäfer and her apprentice Anke Botthof for working hard on the expression and isolation of *Streptomyces* adenylation domains and for their support during the entire erythrochelin project.

For technical assistance I like to acknowledge Christiane Bomm and Gabriele Schimpff-Weiland.

I would also like to thank my diploma student Vanessa Sigmund, whom I cosupervised, for the pleasant time sharing laboratory 4710.

I thank my internship-students Daniela Gutsch, Helene Unger, Kathrin Pieles, Julius Kögel and Felix Terwesten for their participation in my side-projects.

Dr. Verena Helmetag and Dr. Thomas Knappe are gratefully acknowledged for carefully proofreading this thesis.

I would also like to thank my brother Björn Robbel, my sister Kirsten Robbel and their (future) families as well as all of my friends for the outstanding times in Mommenheim.

Most of all, I am indepted to my parents, who have been a continuous financial and moral support throughout my whole life. Their constant motivation and backup in the last years have made all of this possible. It is good to know you have a place you can always return to. For this and so much more I dedicate this thesis to them.